

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 66.1253 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34

Sequence: 1 gactccctcattatcgatcgatcgatgagaca 34

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	34	21	AAA96849
2	19.2	56.5	543	22	AAH65584
3	19.2	56.5	612	22	AAH65585
4	19.2	56.5	666	22	AAH72234
5	19.2	56.5	349980	22	AAH68525
6	19	55.9	447	24	ABH89671
7	19	55.9	521	22	AAH41539
8	19	55.9	521	22	AAH29698
9	19	55.9	2874	22	AAH29578

C	10	19	55.9	3489	22	AAH41061	C	10	19	55.9	3489	22	AAH41061	C	10	19	55.9	3489	22	AAH41061
C	11	18.8	55.3	1015	21	AAH13367	C	11	18.8	55.3	1015	21	AAH13367	C	11	18.8	55.3	1015	21	AAH13367
C	12	18.8	55.3	1884	22	AAH04180	C	12	18.8	55.3	1884	22	AAH04180	C	12	18.8	55.3	1884	22	AAH04180
C	13	18.6	54.7	6075	24	ABH33550	C	13	18.6	54.7	6075	24	ABH33550	C	13	18.6	54.7	6075	24	ABH33550
C	14	18.6	54.7	611590	21	AAH22303	C	14	18.6	54.7	611590	21	AAH22303	C	14	18.6	54.7	611590	21	AAH22303
C	15	18.4	54.1	86	22	AAH23463	C	15	18.4	54.1	86	22	AAH23463	C	15	18.4	54.1	86	22	AAH23463
C	16	18.4	54.1	2601	23	AAH86004	C	16	18.4	54.1	2601	23	AAH86004	C	16	18.4	54.1	2601	23	AAH86004
C	17	18.4	54.1	3188	23	AAH71356	C	17	18.4	54.1	3188	23	AAH71356	C	17	18.4	54.1	3188	23	AAH71356
C	18	18.4	54.1	3336	24	ABH53953	C	18	18.4	54.1	3336	24	ABH53953	C	18	18.4	54.1	3336	24	ABH53953
C	19	18.4	54.1	6987	24	ABH53954	C	19	18.4	54.1	6987	24	ABH53954	C	19	18.4	54.1	6987	24	ABH53954
C	20	18.4	54.1	6990	24	ABH53959	C	20	18.4	54.1	6990	24	ABH53959	C	20	18.4	54.1	6990	24	ABH53959
C	21	18.4	54.1	1163020	24	ABH67197	C	21	18.4	54.1	1163020	24	ABH67197	C	21	18.4	54.1	1163020	24	ABH67197
C	22	18.4	54.1	1163020	24	ABH67197	C	22	18.4	54.1	1163020	24	ABH67197	C	22	18.4	54.1	1163020	24	ABH67197
C	23	18.2	53.5	884	18	AAH59476	C	23	18.2	53.5	884	18	AAH59476	C	23	18.2	53.5	884	18	AAH59476
C	24	18.2	53.5	11601	12	AAH013608	C	24	18.2	53.5	11601	12	AAH013608	C	24	18.2	53.5	11601	12	AAH013608
C	25	18.2	53.5	13058	14	AAH048231	C	25	18.2	53.5	13058	14	AAH048231	C	25	18.2	53.5	13058	14	AAH048231
C	26	18.2	53.5	30032	22	ABH17086	C	26	18.2	53.5	30032	22	ABH17086	C	26	18.2	53.5	30032	22	ABH17086
C	27	18.2	53.5	534720	19	AAH30458	C	27	18.2	53.5	534720	19	AAH30458	C	27	18.2	53.5	534720	19	AAH30458
C	28	18.2	53.5	61890	23	ABH16586	C	28	18.2	53.5	61890	23	ABH16586	C	28	18.2	53.5	61890	23	ABH16586
C	29	18.2	53.5	534720	19	AAH30458	C	29	18.2	53.5	534720	19	AAH30458	C	29	18.2	53.5	534720	19	AAH30458
C	30	18.2	53.5	536165	19	AAH30459	C	30	18.2	53.5	536165	19	AAH30459	C	30	18.2	53.5	536165	19	AAH30459
C	31	18	52.9	565	23	ABH55233	C	31	18	52.9	565	23	ABH55233	C	31	18	52.9	565	23	ABH55233
C	32	18	52.9	4259	21	AAH74870	C	32	18	52.9	4259	21	AAH74870	C	32	18	52.9	4259	21	AAH74870
C	33	18	52.9	5305	20	AAH28289	C	33	18	52.9	5305	20	AAH28289	C	33	18	52.9	5305	20	AAH28289
C	34	18	52.9	5591	23	ABH04293	C	34	18	52.9	5591	23	ABH04293	C	34	18	52.9	5591	23	ABH04293
C	35	18	52.9	13437	23	ABH04292	C	35	18	52.9	13437	23	ABH04292	C	35	18	52.9	13437	23	ABH04292
C	36	18	52.9	1038602	20	AAH01425	C	36	18	52.9	1038602	20	AAH01425	C	36	18	52.9	1038602	20	AAH01425
C	37	18	52.9	1503900	22	AAH95240	C	37	18	52.9	1503900	22	AAH95240	C	37	18	52.9	1503900	22	AAH95240
C	38	18	52.9	1503900	22	AAH95733	C	38	18	52.9	1503900	22	AAH95733	C	38	18	52.9	1503900	22	AAH95733
C	39	17.8	52.4	371	23	AAH65536	C	39	17.8	52.4	371	23	AAH65536	C	39	17.8	52.4	371	23	AAH65536
C	40	17.8	52.4	1804	22	AAH09226	C	40	17.8	52.4	1804	22	AAH09226	C	40	17.8	52.4	1804	22	AAH09226
C	41	17.8	52.4	1804	22	AAH7850	C	41	17.8	52.4	1804	22	AAH7850	C	41	17.8	52.4	1804	22	AAH7850
C	42	17.8	52.4	1804	22	AAH85124	C	42	17.8	52.4	1804	22	AAH85124	C	42	17.8	52.4	1804	22	AAH85124
C	43	17.8	52.4	3371	22	AAH21261	C	43	17.8	52.4	3371	22	AAH21261	C	43	17.8	52.4	3371	22	AAH21261
C	44	17.8	52.4	1082138	21	AAH22305	C	44	17.8	52.4	1082138	21	AAH22305	C	44	17.8	52.4	1082138	21	AAH22305
C	45	17.6	51.8	507	24	ABH30419	C	45	17.6	51.8	507	24	ABH30419	C	45	17.6	51.8	507	24	ABH30419

ALIGNMENTS

RESULT 1
AAA96849
ID AAA96849 standard; DNA; 34 BP.
XX
AC AAA96849;
XX
DT 19-FEB-2001 (first entry)
XX
DE Guide desoxynucleotide building block G1.
XX
KW Promoter; intergenic region; Comelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; ss.
XX
OS Synthetic.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99PR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises

PT Sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PS
XX
XX Disclosure; Page 24; 91pp; English.
XX
XX The present sequence represents a guide desoxynucleotide building
CC block, which was used to link directional desoxynucleotide building
CC blocks during construction of chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commaless yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 34 BP; 8 A; 9 C; 7 G; 10 T; 0 other;
SQ
Query Match 100.0%; Score 34; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 3.5e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 GACTCCTTACTTATCGATCGGTACTGTGAGACA 34
DB 1 GACTCCTTACTTATCGATCGGTACTGTGAGACA 34
RESULT 2
AAH65584
ID AAH65584 standard; DNA; 543 BP.
AC AAH65584;
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 619.
DE
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tereishi N, Senoh A, Ikeda M, Ozaki A;
PI
XX
XX WPI; 2001-376931/40.
DR
XX
XX P-PSDB; AAG90365.
PT
XX
XX Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX Claim 8; SEQ ID NO: 619; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum Corynebacterium glutamicum. These

are useful for identifying the mutation point of a gene derived from a mutant of *Corynebacterium* bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from *Corynebacterium* bacterium, and identifying a homologue of a gene derived from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Seq Sequence 543 BP; 112 A; 160 C; 145 G; 126 T; 0 other;

Query Match 56.5%; Score 19.2; DB 22; Length 543;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 GACTCCTCTACTTATCCGATCGGTACTCTGGAGA 32
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 GACTGCCTCTCTGATCGATCACTACGTAGACA 472

RESULT 3
AAH65585/c
ID AAH65585 standard; DNA; 612 BP.
XX AAH65585;
XX
XX
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 620.
XX
XX
KW *Coryneform bacterium*; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
XX
OS *Corynebacterium glutamicum*.
FN EP1108790-A2.
PN
XX
PD 20-JUN-2001.
PR
XX
PE 18-DEC-2000; 2000EP-0127688.
PP
XX
PI 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX
PY (KYOW) KYOMA HAKKO KOGYO KK.
PA
XX
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX
WP1: 2001-376931/40.
DR P-FSDB; AMG90366.
XX
XX
PT Novel polynucleotides derived from *Coryneform* bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX
PS Claim 8; SEQ ID NO: 620; 246bp + Sequence Listing; English.
XX
XX
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the *Coryneform* bacterium *Corynebacterium glutamicum*. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of *Coryneform* bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from *Coryneform* bacterium, and identifying a homologue of a gene derived
CC from *Coryneform* bacterium. *Coryneform* bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

SO Sequence 612 BP; 147 A; 159 C; 171 G; 135 T; 0 other;

Query Match 56.5%; Score 19.2; DB 22; Length 612;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAGA 32
 DB 154 GACTGCTCCTCTATCATCATCACTACGCTGACA 123

RESULT 4
 ID AAF72234 standard; DNA; 666 BP.

AC AAF72234;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:963.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

OS Corynebacterium glutamicum.

PN WO200100843-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-1B00923.

XX 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.

PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041384.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.

PA (BADI) BASF AG.

PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habernauer G;

DR WPI; 2001-137957/14.

XX P-PADB; AAB80115.

PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteogenic amino acids,
 PT and purine and pyrimidine bases -
 XX Claim 3; Page 1476; 1737p; English.

CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals.
 CC in microorganisms, including organic acids, nonproteogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.

XX Sequence 666 BP; 147 A; 189 C; 171 G; 159 T; 0 other;

Query Match 56.5%; Score 19.2; DB 22; Length 666;
 Best Local Similarity 75.0%; Pred. No. 40;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAGA 32
 DB 541 GACTGCTCCTCTATCATCATCACTACGCTGACA 572

RESULT 5

ID AAH68525 standard; DNA; 349980 BP.

XX AAH68525;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 7060.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

XX

PN EP108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159152.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
 XX
 DR WPI; 2001-376931/40.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Disclosure; SEQ ID NO: 7060; 246bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;
 XX
 Query Match 56.5%; Score 19.2; DB 2; Length 349980;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 XX
 QY 1 GACTCCTACTTATCGATCGTACGTGAGA 32
 Db 271214 GACTGCTCTCTGATCGATCGTACGTGAGA 271245
 XX
 RESULT 6
 ABL89671/c
 ID ABL89671 standard; cDNA; 447 BP.
 XX
 AC ABL89671;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 233.
 XX
 KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US16450.
 XX

PR 19-MAY-2000; 2000US-20551SP.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-122018/16.
 DR P-FSDB; ABB89262.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 XX
 PS Claim 4; SEQ ID NO 233; 2081bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 447 BP; 105 A; 120 C; 120 G; 101 T; 1 other;
 XX
 Query Match 55.9%; Score 19; DB 24; Length 447;
 Best Local Similarity 81.5%; Pred. No. 46;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 XX
 QY 1 GACTCCTCTTATCGATCGTACTG 27
 Db 173 GACTGCCCTGCTTATCGATCGTACTG 147
 XX
 RESULT 7
 AAS41599/c
 ID AAS41599 standard; cDNA; 521 BP.
 XX
 AC AAS41599;
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human enzyme polypeptide #815.
 XX
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; neurotropic; anticoagulant; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155301-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 XX

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAR-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PT Rosen CA, Barash SC, Ruben SM;
XX
XX
DR WPI; 2001-465566/50.
XX
XX P-PSDB; AAU23729.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT Preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases

XX Claim 4; SEQ ID No 825; 1180bp; English.

XX The present invention relates to the isolation of novel human enzyme

CC polypeptides (AAU22915-AAU23814), and the CDNA and genomic sequences

CC encoding them. The enzyme polypeptides of the invention may comprise the

CC functional classes of oxidoreductases, transferases, hydrolases, lyases,

CC isomerases or ligases. The sequences of the invention are useful in the

CC diagnosis, treatment, prevention and/or prognosis of a wide range of

CC disorders including hyperproliferative disorders (e.g. cancer),

CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders

CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),

CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders

CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),

CC blood-related disorders (e.g. haemophilia), reproductive disorders

CC (e.g. infertility) and infectious disorders (e.g. influenza). The

CC polynucleotides of the invention can also be used in gene therapy.

CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human

CC enzyme polypeptides of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 521 BP; 120 A; 145 C; 135 G; 116 T; 5 other;

Query Match 55.9%; Score 19; DB 22; Length 521;

Best Local Similarity 81.5%; Pred. No. 48;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GACTGCTTACTTATGATCGGACTG 27

Db 198 GACTGCGCTGCTTATGATCGGACTG 172

RESULT 8

AAU29698/C

ID AAU29698 standard; cDNA; 521 BP.

XX

AC AAU29698;

XX

DT 21-NOV-2001 (first entry)

XX

DE Human endocrine polypeptide encoding cDNA SEQ ID No 198.

XX

KM Endocrine protein; human; mouse; rabbit; goat; horse; food additive;

KM cat; dog; chicken; sheep; immunosuppressive; antidiabetic; vasotropic;

KM antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;

KM cerebroprotective; neurotropic; antibacterial; virucide; fungicide; cancer;

KM ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;

KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;

KM cerebrovascular disorder; nervous system disorder; bacterial infection;

KM fungal infection; viral infection; ocular disorder; endocrine disorder;

KM gastrointestinal disorder; renal disorder; respiratory disorder;

KM wound healing; skin aging; organ transplantation; food preservative;

KM tissue regeneration; anti-infertility.

XX

OS Homo sapiens.

XX

PN M0200155364-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01308.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189076.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216680.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 14-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225477.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226688.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234977.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246513.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451936/48.
DR P-PSDB; AAU18469.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders of the endocrine system such as reproductive disorders,
XX endocrine cancers and also for testing and detection e.g. diagnosis -
XX
XX Claim 1, SEQ ID No 198; 604bp; English.
XX
CC Sequences AAS29578-1-AAS29736 represent cDNA molecules, which encode the
CC endocrine polypeptides of the invention. Endocrine polypeptides and their
CC associated polynucleotides of the invention are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A

CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 55.9%; Score 19; DB 22; Length 521;
Best Local Similarity 81.5%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GACTCCTCTACTTATCATGCTACTG 27
Db 198 GACTGCGCTGCTTATGATCATGCTACTG 172

RESULT 9
AAS29578/c
ID AAS29578 standard; cDNA; 2874 BP.
XX
AC AAS29578;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human endocrine polypeptide encoding cDNA SEQ ID No 78.
XX
XX Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasoprotic;
KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nocotropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerability; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility.
XX
OS Homo sapiens.
XX
PN NO200155364-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01308.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0214886.
PR 28-JUN-2000; 2000US-0215135.
PR 30-JUN-2000; 2000US-0216680.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

[illegible]

PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251088
PR	06-DEC-2000	2000US-02516719
PR	06-DEC-2000	2000US-0251679
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251890
PR	08-DEC-2000	2000US-0251989
PR	11-DEC-2000	2000US-0254097
PR	05-JAN-2001	2000US-0259678
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM,	
XX	WPI, 2001-451936/48.	
DR	P-PSDB; AAU18349.	
XX		
PT	Isolated polypeptide for treating	
PT	disorders of the endocrine system	
PT	endocrine cancers and also for t	
XX		
PS	Claim 1; SEQ ID NO 78; 604bp; E	
XX		
CC	Sequences AAS29511-AAS29736 rep	
CC	associate polynucleotides of the	
CC	treatment and prevention of vari	
CC	mice, rabbits, goats, horses, ca	
CC	pathological condition can be de	
CC	absence of a mutation in an endo	
CC	disorders include autoimmune dis	
CC	hyperproliferative disorders such	
CC	cardiovascular disorders such as	
CC	disorders such as cerebral ischa	
CC	Alzheimer's disease, infections	

Isolated polypeptide for treating, preventing and/or prognosing disorders of the endocrine system such as reproductive disorders, endocrine cancers and also for testing and detection e.g. diagnosis -

Sequences MA29511-MA29726 represent cDNA molecules, which encode the endocrine polypeptides of the invention. Endocrine polypeptides and their associated polynucleotides of the invention are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by determining the presence or absence of a mutation in an endocrine polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi.

CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptide can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 55.9%; Score 19; DB 22; Length 2874;

Best Local Similarity 81.5%; Pred. No. 62;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GACTGCTTACTTATCGATCGTACTG 27
Db 198 GACTGCTTACTTATCGATCGTACTG 172

RESULT 10

AA541061/c
ID AA541061 standard; cDNA; 3489 BP.

XX AA541061;

DT 17-DEC-2001 (first entry)

XX cDNA encoding novel human enzyme polypeptide #277.

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.

XX Homo sapiens.

PN W0200155301-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01239.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217467.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232402.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234474.
PR 25-SEP-2000; 2000US-0234897.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.

Query Match 55.3%; Score 18.8; DB 21; Length 1015;
Best Local Similarity 76.7%; Pred. No. 65;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACTCCTACTTATGATCGGCTACTGTGAG 31
DB 439 AGTCCTACTTATGATCGGCTACTGTG 468

RESULT 12

ADD04180
ID ADD04180 standard; DNA; 1884 BP.

XX AAD04180;

DT 02-JUL-2001 (first entry)

XX Aspergillus oryzae lysophospholipase 1 (LPL-1) DNA.

XX lysophospholipase; LPL-1; lysophospholipid; lecithin; filterability;
XX bakery product; dough; bread; cake; elasticity; edible oil; rape seed;
XX soybean; sunflower; ds.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT CDS 1..1884

FT /tag= a /product= "Aspergillus oryzae LPL-1 protein"

FT sig_peptide 1..63

FT /tag= b 64..1881

FT mat_peptide

FT /tag= c

FT /product= "Mature A. oryzae LPL-1 protein"

XX WO200127251-A1.

XX 19-APR-2001.

XX 12-OCT-2000; 2000WO-DK00577.

XX 14-OCT-1999; 99DK-0001473.

XX (NOVO) NOVOTYMES AS.

XX Udagawa H, Frandsen TP, Nielsen TAB, Kaupinen MS, Christensen S;

XX WPI; 2001-282018/29.

XX P-PSDB; AAE00814.

XX New lysophospholipase enzyme useful in hydrolyzing fatty acyl groups in
XX an aqueous solution or slurry of carbohydrate origin -

XX Claim 4; Page 39-42; 58pp; English.

XX The present sequence is a DNA encoding Aspergillus oryzae

XX lysophospholipase-1 (LPL-1). Lysophospholipase is an enzyme that

XX hydrolyses lysophospholipids into fatty acids. The DNA encoding LPL-1

XX is cloned into a plasmid present in Escherichia coli deposit number

XX DSM 13082. Lysophospholipase is used to hydrolyse fatty acyl groups

XX in a phospholipid or lysophospholipid and for improving filterability

XX of an aqueous solution or slurry of carbohydrate origin which contains

XX phospholipid and preferably a starch hydrolysate, particularly wheat

XX starch hydrolysate. Lysophospholipase is also used to release fatty

XX acids from intact phospholipid e.g. lecithin, and in the preparation

XX of bakery products such as dough, bread, and cakes to improve elasticity.

XX Lysophospholipase is used to reduce the phospholipid content in an

XX edible oil e.g. vegetable oils such as rape seed oil, soybean oil and

XX sunflower oil.

XX Sequence 1884 BP; 429 A; 556 C; 463 G; 436 T; 0 other;

Query Match 55.3%; Score 18.8; DB 22; Length 1884;
Best Local Similarity 76.7%; Pred. No. 72;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACTCCTACTTATGATCGGCTACTGTGAG 31
DB 425 AGTCCTACTTATGATCGGCTACTGTG 454

RESULT 13

ABL33550/c
ID ABL33550 standard; DNA; 6075 BP.

XX ABL33550;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1523.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antiamebic; cyrostatic; noctropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antineumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

XX cytosine methylation -

XX Claim 1; SEQ ID NO 1523; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

XX genes which are modified by the methylation of cytosines. The sequences

XX can be used in the diagnosis and treatment of immune system disorders,

XX including eye diseases such as retinopathy, neovascular glaucoma and

XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

XX diseases. The present sequence is a gene of the invention.

XX Sequence 6075 BP; 1318 A; 146 C; 1645 G; 2966 T; 0 other;

XX Query Match 54.7%; Score 18.6; DB 24; Length 6075;

XX Best Local Similarity 72.7%; Pred. No. 1.1e+02;

XX Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

XX QY 2 ACTCCTACTTATGATCGGCTACTGTGAGCA 34

XX DB 859 ACACCTACTTATGATCGGCTACTGTG 827

XX RESULT 14

AAF22303
 ID AAF22303 standard; DNA; 611590 BP.
 XX
 AC AAF22303;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 DE Arabidopsis thaliana chromosome 2 centromere.
 XX
 KM Centromere; microsome; vector; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO20055325-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 17-MAR-2000; 2000WO-US07392.
 XX
 PR 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0133584.
 PR 17-SEP-1999; 99US-0154603.
 XX
 PA (UYCH-) UNIV CHICAGO.
 XX
 PI Preuss D, Copenhaver G, Keith K;
 DR WPI; 2000-587529/55.
 XX
 PT Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells -
 XX
 PS Claim 45; Page 820-959; 1449pp; English.
 XX
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX
 SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;
 Query Match 54.7%; Score 18.6; DB 21; Length 611590;
 Best Local Similarity 72.7%; Pred. No. 2.1e+02;
 Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 GACTCCTCTACTTATCGATCGTACTGTGAGC 33
 Db 384616 GAGACATCTACTTACTTATCGTACTGTGAGC 384648

RESULT 15
 AAS23463/c
 ID AAS23463 standard; DNA; 86 BP.
 XX
 AC AAS23463;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE C. albicans essential gene knockout upstream PCR primer (KO-Up) #21.
 XX
 KM Gene identification; essential gene; GRACE; pathogenic fungus;
 KM gene replacement and conditional expression; fungal infection;
 KM PCR primer; KO-Up; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200160975-A2.

PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05551.
 XX
 PR 18-FEB-2000; 2000US-0183534.
 XX
 PR (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H;
 DR WPI; 2001-489080/53.
 XX
 PT Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes -
 XX
 PS Disclosure; Page 255; 324pp; English.
 XX
 CC The present invention relates to novel methods for constructing fungal
 CC strains useful for identification and validation of gene products as
 CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs.
 CC The invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. Candida albicans,
 CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
 CC useful to identify agents that may be used in the treatment of fungal
 CC infections. AAS23463-AAS23503 represent C. albicans essential gene
 CC knockout upstream PCR primers (KO-Up) used in the methods of the
 CC present invention.
 XX
 SQ Sequence 86 BP; 28 A; 20 C; 10 G; 28 T; 0 other;
 Query Match 54.1%; Score 18.4; DB 22; Length 86;
 Best Local Similarity 78.6%; Pred. No. 68;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 4 TCCCTACTTATCGATCGTACTGTGAG 31
 Db 83 TCCACTAGTTCTAGAGCGTTCTGTGAG 56

Search completed: May 11, 2003, 03:06:06
 Job time: 188.125 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 13.6341 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34
Sequence: 1 gactcctcattcgcacgcgtgagaca 34

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.4	54.1	3336	4	US-09-026-033-1	Sequence 1, Appli
C 2	18.4	54.1	3336	4	US-09-026-033-2	Sequence 2, Appli
C 3	18.4	54.1	6987	4	US-09-026-033-3	Sequence 3, Appli
C 4	18.4	54.1	6990	4	US-09-026-033-23	Sequence 23, Appli
C 5	18.2	53.5	11601	2	US-08-222-617A-3	Sequence 24, Appli
C 6	18.2	53.5	11601	2	US-08-222-617A-24	Sequence 24, Appli
C 7	17.6	51.8	507	4	US-09-280-116-189	Sequence 189, App
C 8	17.6	51.8	933	4	US-09-280-116-191	Sequence 191, App
C 9	17.4	51.2	2432	4	US-08-974-022-1	Sequence 1, Appli
C 10	17.4	51.2	2432	4	US-08-795-445A-1	Sequence 1, Appli
C 11	17.4	51.2	2432	4	US-08-974-186-1	Sequence 1, Appli
C 12	17.4	51.2	2432	4	US-08-795-446B-1	Sequence 1, Appli
C 13	17.4	51.2	2432	4	US-08-795-446B-1	Sequence 1, Appli
C 14	17.4	51.2	2432	4	US-08-706-945D-123	Sequence 123, App
C 15	17.2	50.6	993	1	US-08-364-081-2	Sequence 2, Appli
C 16	17.2	50.6	993	1	US-08-630-552-2	Sequence 2, Appli
C 17	17.2	50.6	993	5	PCT-US95-16558-2	Sequence 2, Appli
C 18	17.2	50.6	1011	1	US-08-375-235-1	Sequence 1, Appli
C 19	17.2	50.6	1183	1	US-07-755-009-1	Sequence 1, Appli
C 20	17.2	50.6	1183	1	US-07-755-009-2	Sequence 1, Appli
C 21	17.2	50.6	1860	2	US-08-331-644-3	Sequence 3, Appli
C 22	17.2	50.6	1860	5	PCT-US93-04102-3	Sequence 3, Appli
C 23	17.2	50.6	4910	4	US-09-303-064-40	Sequence 40, Appli
C 24	17.2	50.6	4910	4	US-09-086-503-40	Sequence 40, Appli
C 25	17.2	50.0	249	4	US-09-134-001C-1324	Sequence 1324, Ap
C 26	17.2	50.0	1303	4	US-08-894-440-2	Sequence 2, Appli
C 27	17.2	50.0	1303	4	US-09-458-093-2	Sequence 2, Appli

C 28	17	50.0	17710	4	US-08-976-259-70	Sequence 70, Appli
C 29	16.8	49.4	240	1	US-08-054-985-1	Sequence 1, Appli
C 30	16.8	49.4	1177	3	US-08-611-587-7	Sequence 7, Appli
C 31	16.8	49.4	1240	4	US-09-221-017B-844	Sequence 844, App
C 32	16.6	48.8	500	2	US-08-967-101-50	Sequence 50, Appli
C 33	16.6	48.8	500	2	US-08-592-541-50	Sequence 50, Appli
C 34	16.6	48.8	500	3	US-09-124-698-50	Sequence 50, Appli
C 35	16.6	48.8	500	3	US-09-124-698-50	Sequence 50, Appli
C 36	16.6	48.8	500	4	US-08-496-841C-50	Sequence 50, Appli
C 37	16.6	48.8	500	4	US-09-124-523-50	Sequence 50, Appli
C 38	16.6	48.8	611	4	US-09-221-017B-352	Sequence 352, App
C 39	16.6	48.8	736	2	US-08-967-101-12	Sequence 12, Appli
C 40	16.6	48.8	736	2	US-08-967-101-155	Sequence 155, App
C 41	16.6	48.8	736	2	US-08-592-541-12	Sequence 12, Appli
C 42	16.6	48.8	736	2	US-08-592-541-155	Sequence 155, App
C 43	16.6	48.8	736	3	US-08-888-077A-15	Sequence 15, Appli
C 44	16.6	48.8	736	3	US-09-124-698-12	Sequence 12, Appli
C 45	16.6	48.8	736	3	US-09-124-698-155	Sequence 155, App

ALIGNMENTS

RESULT 1
US-09-026-033-1/c
Sequence 1, Application US/09026033
Patent No. 6368791
GENERAL INFORMATION:
APPLICANT: Felix, Carolyn A.
APPLICANT: Jones, Douglas H.
APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
TITLE OF INVENTION: CHROMOSOMAL REARRANGEMENTS ASSOCIATED WITH LEUKEMIA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-1001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3336 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
TITLE OF INVENTION: CHROMOSOMAL REARRANGEMENTS ASSOCIATED WITH LEUKEMIA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
FLOOR
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-10011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 6990 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-033-23

Query Match 54.1%; Score 18.4; DB 4; Length 6990;
Best Local Similarity 78.6%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 TCTACTATGATCGGTAAGTGTGACA 34
DB 6196 TTACTATGATGATGCGACTGTGGCCA 6223

RESULT 5
US-08-222-617A-3
Sequence 3, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan P.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Montenegro, Eduardo P.
APPLICANT: Van Liempt, Henk
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 11601 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Acetomonium chrysogenum
FEATURE:
NAME/KEY: CDS
LOCATION: 388..11526
OTHER INFORMATION: /function= "Enzyme"
OTHER INFORMATION: /product= "ACV Synthetase"
OTHER INFORMATION: /notes= "NNN=GCC, AGU, AGC, UCU, UCC, UCA, or UCG;"
US-08-222-617A-3

Query Match 53.5%; Score 18.2; DB 2; Length 11601;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 ACTATGATCGGTAAGTGTGACA 32
DB 11163 ACTATGATCGGTAAGTGTGACA 11185

RESULT 6
US-08-222-617A-24
Sequence 24, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan P.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Montenegro, Eduardo P.
APPLICANT: Van Liempt, Henk
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION: 97,157
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 11601 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Acremonium chrysogenum
FEATURE:
NAME/KEY: CDS
LOCATION: 388..11526
OTHER INFORMATION: /function= "Enzyme"
OTHER INFORMATION: /product= "Acv Synthetase"
FEATURE:
NAME/KEY: CDS
LOCATION: 8050..8052
OTHER INFORMATION:
US-08-222-617A-24 /note= "NNN=AGU, AGC, UCU, UCC, UCA, or UCG"

Query Match 53.5%; Score 18.2; DB 2; Length 11601;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 ACTTATGATCGTACTGTGAGA 32
Db 11163 ACTTATGATCGTACTGTGACA 11185

RESULT 7
US-09-280-116-189
Sequence 189, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 189
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: astacin/m 12a metalloproteases
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(507)
OTHER INFORMATION: n = a, t, c or g
US-09-280-116-189

Query Match 51.8%; Score 17.6; DB 4; Length 507;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 CTCTACTTATGATCGGTACTGTG 29
Db 11163 CTCTACTTATGATCGGTACTGTG 29

Db 44 CTCTCTTATAGATCGTACTGTG 67
RESULT 8
US-09-280-116-191
Sequence 191, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 191
LENGTH: 933
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: astacin/m 12a metalloproteases
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(933)
OTHER INFORMATION: n = a, t, c or g
US-09-280-116-191

Query Match 51.8%; Score 17.6; DB 4; Length 933;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 CTCTACTTATGATCGGTACTGTG 29
Db 289 CTCTCTTATAGATCGTACTGTG 312

RESULT 9
US-08-974-022-1
Sequence 1, Application US/08974022
Patent No. 6015938
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,022
FILING DATE: 12-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-974-022-1

Query Match 51.2%; Score 17.4; DB 3; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
Db 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 10
US-08-795-445A-1
Sequence 1, Application US/08795445A
Patent No. 6284485
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,445A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-795-445A-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
Db 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 11
US-08-795-447A-1

Sequence 1, Application US/08795447A
Patent No. 6284728
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: One Amgen Center Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91362-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,447A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378D2
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-795-447A-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
Db 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 12
US-08-974-186-1
Sequence 1, Application US/08974186
Patent No. 6284740
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,186
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-974-186-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTTATCGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGATCTCAGA 1628

RESULT 13
US-08-795-446B-1
Sequence 1, Application US/08795446B
Patent No. 6288032
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Denavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326

US-08-795-446B-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTTATCGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGATCTCAGA 1628

RESULT 14
US-08-706-945D-123
Sequence 123, Application US/08706945D
Patent No. 6369027
GENERAL INFORMATION:
APPLICANT: Boyle, William
APPLICANT: Lacey, David
APPLICANT: Calzone, Frank
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
FILE REFERENCE: A-378CIP
CURRENT APPLICATION NUMBER: US/08/706,945D
CURRENT FILING DATE: 1996-09-03
PRIOR APPLICATION NUMBER: 08/577,788
PRIOR FILING DATE: 1995-12-22
NUMBER OF SEQ ID NOS: 145
SOFTWARE: PatentIn version 3.1
SEQ ID NO 123
LENGTH: 2432
TYPE: DNA
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: (124)..(1326)
OTHER INFORMATION:
US-08-706-945D-123

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTTATCGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGATCTCAGA 1628

RESULT 15
US-08-364-081-2/c
Sequence 2, Application US/08364081
Patent No. 5585464
GENERAL INFORMATION:
APPLICANT: Ramesh K. Prakash
TITLE OF INVENTION: Recombinant Antigen for Diagnosing
TITLE OF INVENTION: Rheumatoid Arthritis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thorpe, No. 5585464th & Western
STREET: 9035 South 700 East, Suite 200
CITY: Sandy
STATE: Utah
COUNTRY: USA
ZIP: 84070
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 kb storage
COMPUTER: AST Advantage NB-SX20
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/364,081
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/019,780
 FILING DATE: 19-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Alan J. Howarth
 REGISTRATION NUMBER: 36,553
 REFERENCE/DOCKET NUMBER: T781CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (801)566-6633
 TELEFAX: (801)566-0750
 INFORMATION FOR SEO ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 993 nucleotides
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-364-081-2

Query Match 50.6%; Score 17.2; DB 1; Length 993;
 Best Local Similarity 86.4%; Pred. No. 32;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 10 ACTTATCGATCGCTACTGTGAG 31
 Db 265 ACTTCTCGATTGGAACCTGTGAG 244

Search completed: May 11, 2003, 03:07:47
 Job time: 18.6341 secs


```
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 619
LENGTH: 543
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-619
```

```
Query Match 56.5%; Score 19.2; DB 9; Length 543;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32
DB 441 GACTGCTCTCTGATCGATCACTACGTGAGA 472
```

```
RESULT 3
US-09-738-626-620/c
Sequence 620, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT* SENOH, AKIHIRO
APPLICANT* IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 620
LENGTH: 612
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-620
```

```
Query Match 56.5%; Score 19.2; DB 9; Length 612;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32
```

```
DB 154 GACTGCTCTCTGATCGATCACTACGTGAGA 123
```

```
RESULT 4
US-09-738-626-1
Sequence 62, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
```

```
Query Match 56.5%; Score 19.2; DB 9; Length 3309400;
Best Local Similarity 75.0%; Pred. No. 1,1e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32
DB 571214 GACTGCTCTCTGATCGATCACTACGTGAGA 571245
```

```
RESULT 5
US-10-039-785-62/c
Sequence 62, Application US/10039785
Patent No. US20020067646A1
GENERAL INFORMATION:
APPLICANT: Salcedo et al.
TITLE OF INVENTION: Antibodies that Immunospecifically Bind to TRAIL
FILE REFERENCE: PF550
CURRENT APPLICATION NUMBER: US/10/039,785
CURRENT FILING DATE: 2002-05-07
PRIOR APPLICATION NUMBER: 60/369,860
PRIOR FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: 60/341,237
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/331,310
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/331,044
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: 60/327,364
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/323,807
PRIOR FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: 60/309,176
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/294,981
PRIOR FILING DATE: 2001-06-04
```

;; PRIOR APPLICATION NUMBER: 60/293,473
;; PRIOR FILING DATE: 2001-05-25
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 62
;; LENGTH: 750
;; TYPE: DNA
;; ORGANISM: Artificial sequence
;; FEATURE:
;; OTHER INFORMATION: DNA encoding T1015A02 scfv
US-10-039-785-62

Query Match 54.7%; Score 18.6; DB 12; Length 750;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGGTACTGTGAG 33
DB 235 GATTCTGAGAGTATCCATCATCATGTAAGTC 203

RESULT 6

US-09-783-590-294
;; Sequence 294, Application US/09783590
;; Patent No. US20020110850A1
;; GENERAL INFORMATION:
;; APPLICANT: Dillon, Patrick J.
;; APPLICANT: Haseltine, William A.
;; APPLICANT: Li, Haodong
;; APPLICANT: Rosen, Craig A.
;; APPLICANT: Ruben, Steven M.
;; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;; FILE REFERENCE: PO-16.2C1
;; CURRENT APPLICATION NUMBER: US/09/783,590
;; CURRENT FILING DATE: 2000-02-15
;; PRIOR APPLICATION NUMBER: 08/420,856
;; PRIOR FILING DATE: 1995-04-12
;; PRIOR APPLICATION NUMBER: 08/346,731
;; PRIOR FILING DATE: 1994-11-21
;; NUMBER OF SEQ ID NOS: 12485
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 494
;; LENGTH: 403
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (10)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (141)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (327)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (400)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-294

Query Match 53.5%; Score 18.2; DB 10; Length 403;
Best Local Similarity 74.2%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGGTACTGTGAG 31
DB 337 GAATCTGAACTTTTCAAGGGGTACTGTGAG 367

RESULT 7

US-09-939-964-1/C
;; Sequence 1, Application US/09939964
;; Publication No. US20030054522A1

;; GENERAL INFORMATION:
;; APPLICANT: Rosenthal, Andre
;; APPLICANT: Freiberg, Christoph
;; APPLICANT: Perret, Xavier Philippe
;; APPLICANT: Broughton, William John
;; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
;; FILE REFERENCE: CARP0068
;; CURRENT APPLICATION NUMBER: US/09/939,964
;; CURRENT FILING DATE: 2001-08-27
;; PRIOR APPLICATION NUMBER: 09/214,808
;; PRIOR FILING DATE: 1999-06-22
;; NUMBER OF SEQ ID NOS: 1
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 536165
;; TYPE: DNA
;; ORGANISM: Rhizobium
US-09-939-964-1

Query Match 53.5%; Score 18.2; DB 9; Length 536165;
Best Local Similarity 74.2%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGGTACTGTGAG 31
DB 14804 GACTCCTCTGTTTTCATCGAGCGCTTAG 14774

RESULT 8

US-08-927-939-30/C
;; Sequence 30, Application US/08927939
;; Patent No. US20010006640A1
;; GENERAL INFORMATION:
;; APPLICANT: Grainger, David J.
;; APPLICANT: Tacalick, Lauren Marie
;; TITLE OF INVENTION: Compounds and methods to inhibit or
;; FILE REFERENCE: 295.022US1
;; CURRENT APPLICATION NUMBER: US/08/927,939
;; CURRENT FILING DATE: 1997-09-11
;; NUMBER OF SEQ ID NOS: 83
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 30
;; LENGTH: 4259
;; TYPE: DNA
;; ORGANISM: Mus musculus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (2070)...(2130)
;; NAME/KEY: CDS
;; LOCATION: (2669)...(2795)
;; NAME/KEY: CDS
;; LOCATION: (2990)...(3079)
;; NAME/KEY: CDS
;; LOCATION: (3491)...(3506)
US-08-927-939-30

Query Match 52.9%; Score 18; DB 8; Length 4259;
Best Local Similarity 70.6%; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGGTACTGTGAG 34
DB 1165 GAACCTCTACCCACCAATGGAATGGGAACA 1132

RESULT 9

US-09-946-807-1/C
;; Sequence 1, Application US/09946807
;; Patent No. US20020165144A1
;; GENERAL INFORMATION:
;; APPLICANT: Stefansson, Hreinn

```
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: r=g or a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: y=c/u or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: k=g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: w=a or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: b=g or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: d=a or g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: v=a or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: h=a or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: n=a or g or c or t/u
; US-09-946-807-1
```

```
Query Match 52.9%; Score 18; DB 9; Length 1503841;
Best Local Similarity 70.6%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
RESULT 10
us-09-795-668-1/c
```

```
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: y=c/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: n=a or g or c or t/u
; US-09-795-668-1
```

```
Query Match 52.9%; Score 18; DB 10; Length 1503841;
Best Local Similarity 70.6%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Db 183456 GAATTCCTACTTATGCATCGTACTGTGAGACA 34
183456 GAATTCCTACTTATGCATCGTACTGTGAGAGAAA 183423
```

```
RESULT 11
US-09-795-668-1/c
; Sequence 1, Application US/09795668
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
```

```
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: PatsSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: y=t/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)..(1531)
; OTHER INFORMATION: n=a or g or c or t/u
; US-09-795-686-1

Query Match          52.9%; Score 18; DB 10; Length 1503841;
Best Local Similarity 70.6%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
OY 1 GACTCCTACTTATCGATCGTACTGTGAGACA 34
Db 183456 GAAITTCCTATTATTAATGTGTAAGGAGAAA 183423

RESULT 12
; US-09-918-995-30042
; Sequence 30042, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: PatsSeq for Windows Version 3.0
; SEQ ID NO 30042
; LENGTH: 371
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; US-09-918-995-30042
; Query Match          52.4%; Score 17.8; DB 9; Length 371;
; Best Local Similarity 75.9%; Pred. No. 89;
; Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 5 CCTCTACTTATCGATCGTACTGTGAGAC 33
Db 191 CCTCTACTTATCGATCGTACTGTGAGAC 219

RESULT 13
; US-10-028-072-35/c
; Sequence 35, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
```

PRIOR FILING DATE: 1997-10-31
 PRIOR APPLICATION NUMBER: 60/063127
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063327
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063329
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063550
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063561
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063704
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063733
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063735
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063738
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063755
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064248
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/064809
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: 60/065186
 PRIOR FILING DATE: 1997-11-12
 PRIOR APPLICATION NUMBER: 60/065846
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/066453
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066511
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066770
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/069212
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069278
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069334
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069694
 PRIOR FILING DATE: 1997-12-16
 PRIOR APPLICATION NUMBER: 60/072320
 PRIOR FILING DATE: 1998-01-23
 PRIOR APPLICATION NUMBER: 60/073612
 PRIOR FILING DATE: 1998-02-04
 PRIOR APPLICATION NUMBER: 60/074086
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/074092
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25
 PRIOR APPLICATION NUMBER: 60/079663
 PRIOR FILING DATE: 1998-02-27
 PRIOR APPLICATION NUMBER: 60/079728
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/080165
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/081203
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081229
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081695
 PRIOR FILING DATE: 1998-04-14
 PRIOR APPLICATION NUMBER: 60/081817
 PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/081818
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/082999
 PRIOR FILING DATE: 1998-04-24
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/083545
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/084600
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084627
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084637
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/085149
 PRIOR FILING DATE: 1998-05-12
 PRIOR APPLICATION NUMBER: 60/085323
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085338
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085339
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085579
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085697
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085704
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/086414
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/086430
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/087106
 PRIOR FILING DATE: 1998-05-28
 PRIOR APPLICATION NUMBER: 60/088026
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088730
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088741
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088810
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088858
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/089532
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089599
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089907
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089947
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/090349
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090429
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090445
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090538
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090863
 PRIOR FILING DATE: 1998-06-26
 PRIOR APPLICATION NUMBER: 60/091360
 PRIOR FILING DATE: 1998-07-01
 PRIOR APPLICATION NUMBER: 60/091519
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091982
 PRIOR FILING DATE: 1998-07-07

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
 Best Local Similarity 75.9%; Pred. No. 1.4e+02;
 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 CTCTACTTATGATCGGTACTGTGAGACA 34
DB 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

RESULT 14

US-10-121-049-35/c
; Sequence 35, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 35
; LENGTH: 3371
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-121-049-35

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
Best Local Similarity 75.9%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 CTCTACTTATGATCGGTACTGTGAGACA 34
DB 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

RESULT 15

US-10-123-904-35/c
; Sequence 35, Application US/10123904
; Publication No. US2003002238A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 35
; LENGTH: 3371
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-904-35

FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 35
; LENGTH: 3371
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-904-35

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
Best Local Similarity 75.9%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 CTCTACTTATGATCGGTACTGTGAGACA 34
DB 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

Search completed: May 11, 2003, 06:27:57
Job time : 1007.66 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 546.556 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-15
Perfect score: 34
Sequence: 1 gactccctcattcgcgcgtcgtgagaca 34

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estdb:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gse:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pln:*
21: em_gse_vit:*
22: em_gse_fun:*
23: em_gse_mam:*
24: em_gse_mus:*
25: em_gse_other:*
26: em_gse_pro:*
27: em_gse_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.2	62.4	665	17	AZ607800 1M0430M22
2	20.8	61.2	589	13	BM096067 f330c09.x
3	20.4	60.0	599	13	BI672108 f334f03.x
4	20.4	60.0	660	14	BQ109731 imagegc.7
5	20.2	59.4	702	13	BI823138 603039579
6	20.2	58.8	136	12	BG833308 951004F05

C	7	19.8	58.2	349	12	BF542321	BF542321	UI-R-C2P-
C	8	19.8	58.2	451	9	AL387696	AL387696	MEBC44B09
C	9	19.8	58.2	502	13	BM209528	BM209528	CO649B07-
C	10	19.8	58.2	514	14	BO559633	BO559633	H4059D12-
C	11	19.8	58.2	524	10	BB764031	BB764031	BB764031
C	12	19.8	58.2	530	10	BB752196	BB752196	BB752196
C	13	19.8	58.2	556	17	AZ331947	AZ331947	1M060J06
C	14	19.8	58.2	611	17	AZ831659	AZ831659	2M01H02
C	15	19.8	58.2	615	14	BM945933	BM945933	UI-M-EMO-
C	16	19.6	57.6	165	12	BE816979	BE816979	RC2-BN024
C	17	19.6	57.6	373	14	BQ663890	BQ663890	HU04H190
C	18	19.6	57.6	771	12	BF104508	BF104508	BF104508
C	19	19.6	57.6	979	10	BE293067	BE293067	601106072
C	20	19.6	57.6	1002	12	BE875770	BE875770	601487228
C	21	19.4	57.1	234	17	AZ007407	AZ007407	RC1-23-3
C	22	19.4	57.1	259	10	BE593348	BE593348	BM593348
C	23	19.4	57.1	665	14	BQ086932	BQ086932	CF1_10_CO
C	24	19.4	57.1	690	17	BH265000	BH265000	CH230-665
C	25	19.4	57.1	713	13	BI521708	BI521708	603081645
C	26	19.2	56.5	213	10	AM445625	AM445625	81979 MAR
C	27	19.2	56.5	249	9	AL616637	AL616637	zehu0382.
C	28	19.2	56.5	304	17	CNS000P6	AL084576	Arabidops
C	29	19.2	56.5	332	13	BM573163	BM573163	FX96C02.Y
C	30	19.2	56.5	332	13	BM573789	BM573789	FX03H03.Y
C	31	19.2	56.5	374	17	BH348403	BM157390	CH230-184
C	32	19.2	56.5	380	13	BM157390	BM157390	FX42D06.Y
C	33	19.2	56.5	389	13	BM142056	BM142056	FX68A06.X
C	34	19.2	56.5	414	14	BQ073968	BQ073968	FX33801.X
C	35	19.2	56.5	445	13	BM16704	BM16704	FX83506.X
C	36	19.2	56.5	484	14	D32654	D32654	CELEK01BYR
C	37	19.2	56.5	500	17	AQ148607	AQ148607	HS_3137_B
C	38	19.2	56.5	505	10	AW777682	AW777682	FX49B06.X
C	39	19.2	56.5	524	9	AA593744	AA593744	nl83910.S
C	40	19.2	56.5	543	17	AZ289814	AZ289814	RC1-23-4
C	41	19.2	56.5	547	14	BQ133481	BQ133481	FX42807.Y
C	42	19.2	56.5	572	17	AZ440572	AZ440572	1M0231M12
C	43	19.2	56.5	582	17	AQ673369	AQ673369	HS_5486_A
C	44	19.2	56.5	597	10	AW777842	AW777842	FX51E02.X
C	45	19.2	56.5	611	14	BM777458	BM777458	FX31D03.Y

ALIGNMENTS

RESULT 1
LOCUS AZ607800
DEFINITION 665 bp DNA linear GSS 13-DEC-2000
ACCESSION IM0430M22F Mouse 10kb plasmid UGCM library Mus musculus genomic
VERSION AZ607800
KEYWORDS AZ607800.1 GI:11729990
SOURCE GSS.
ORGANISM house mouse.
REFERENCE Mus musculus.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 665)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weis,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 B., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00

Plate: 0430 row: M column: 22
 Seq primer: CATTGTAACACGACGCAGT
 Class: plasmid ends
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES

SOURCE

1. 665
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0010430M22"
 /clone_1ib="Mouse 10kb plasmid U0010430M22"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptor and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

192 a 131 c 115 g 227 t

ORIGIN

Query Match 62.4%; Score 21.2; DB 17; Length 665;
 Best Local Similarity 76.5%; Pred. No. 2.3e+02;
 Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Oy 1 GACTCCTCTACTTATCGATCGTACTGTGAGCA 34
 Db 545 GCCTCCTTTAAATCAATCGATGTGAGCA 578

RESULT 2
 BM096067 589 bp mRNA linear EST 20-NOV-2001
 LOCUS t330c09.x1 Sugano SJD adult male Danio rerio cDNA clone 5409809 3'
 DEFINITION t330c09.x1 Sugano SJD adult male Danio rerio cDNA clone 5409809 3', mRNA sequence.
 ACCESSION BM096067
 VERSION BM096067.1 GI:17025033
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 589)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbratfish@wustl.edu

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

FEATURES

SOURCE

Library constructed by Dr. Sumio Sugano and Dr. Kolchi Kawakami.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI, send email to: info@image.llnl.gov
 zebrafish identity (p-value greater than 1e-99) found to:
 g12446832[gb|AA606179|AA606179.fal6e02.st Appel Eisen zebrafish embryo 15
 High quality sequence stop: 463.
 Location/Qualifiers
 1. 589
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="5409809"
 /clone_1ib="Sugano SJD adult male"
 /sex="male"
 /tissue_type="whole body"
 /dev_stage="adult"
 /lab_host="DH10B (phage resistant)"
 /note="Vector: pME18S-FL3; Site 1: DraIII (CACCATGTG); Site 2: DraIII (CAGGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGACCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TTTGGCCTTCTG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTCCTCTAAAGCTGCG and 3' end primer CGACCTGACGCTGAGCACA."

BASE COUNT

125 a 145 c 172 g 147 t

ORIGIN

Query Match 61.2%; Score 20.8; DB 13; Length 589;
 Best Local Similarity 78.1%; Pred. No. 3.2e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 Oy 1 GACTCCTCTACTTATCGATCGTACTGTGAGA 32
 Db 10 GATTCCTGTCTTATTCATCGCAGCTGTGACA 41

RESULT 3
 B1672108/c 599 bp mRNA linear EST 12-SEP-2001
 LOCUS t334f03.x1 Gong zebrafish testis Danio rerio cDNA clone 5152637 3'
 DEFINITION t334f03.x1 Gong zebrafish testis Danio rerio cDNA clone 5152637 3', mRNA sequence.
 ACCESSION B1672108
 VERSION B1672108.1 GI:15587492
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 599)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbratfish@wustl.edu

REFERENCE

AUTHORS

TITLE
 JOURNAL
 COMMENT

The library was constructed by Dr. Z. Gong. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Please

contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T7 from Gibco
High quality sequence stop: 463.

FEATURES

source

Location/Qualifiers

1. .599

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="5152637"

/clone_1lb="Gong zebrafish testis"

/sex="male"

/dev_stage="4-5 month"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis (pooled); Vector: pBluescript SK-"

Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from

the testes of 31 male adult zebrafish (4-5 month old).

CDNAs were made using oligo-dT primers and inserted into

lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo

mass-excised to pBluescript SK- following the Washington

University protocol

(http://genome.wustl.edu/esf/lambda_protocol.shtml).

Please contact Zhiyuan Gong for further information on

this library (National University of Singapore,

Department of Biological Sciences, Lower Kent Ridge Road,

Singapore 119260)."

BASE COUNT 198 a 82 c 117 g 202 t

ORIGIN

Query Match 60.0%; Score 20.4; DB 13; Length 599;
Best Local Similarity 80.0%; Pred. No. 4.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CTCTCTACTTATGATCGGTAAGTGTGAGA 32

DB 386 CTCTCTCCCTTTTGATCGAAGTGTGAGA 357

RESULT 4
BO109731/c 660 bp mRNA linear EST 16-APR-2002

LOCUS imagec_7_2001/bm420bdf41.xl Gong zebrafish testis Danio rerio

DEFINITION cDNA clone IMAGE:5152637 3', mRNA sequence.

BO109731

BO109731.1 GI:20159385

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

zefrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

1 (bases 1 to 660)

Kale,P.L., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and

Prange,C.K.

The I.M.A.G.E. Consortium quality control effort: clone

resequencing for verification

Unpublished (2001)

Contact: Prange CK

The I.M.A.G.E. Consortium

Lawrence Livermore National Laboratory

Livermore, CA, USA

Email: help@image.llnl.gov

This read has been produced as part of the I.M.A.G.E. Consortium

quality control effort. High quality sequence is defined as having

100 or more base pairs with a phased quality value of 20 or greater,

where a sliding window of 4 base pairs with a phased quality value

of 15 or greater marks the beginning and end of the sequence. For

information on obtaining this clone, please contact

image@llnl.gov.

plate: LLAM11377 row: k column: 6

Seq primer: -21m13

High quality sequence stop: 660.

location/Qualifiers

source

1. .660

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="IMAGE:5152637"

/clone_1lb="Gong zebrafish testis"

/sex="male"

/dev_stage="4-5 month"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis (pooled); Vector: pBluescript SK-"

Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from

the testes of 31 male adult zebrafish (4-5 month old).

CDNAs were made using oligo-dT primers and inserted into

lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo

mass-excised to pBluescript SK- following the Washington

University protocol

(http://genome.wustl.edu/esf/lambda_protocol.shtml).

Please contact Zhiyuan Gong for further information on

this library (National University of Singapore,

Department of Biological Sciences, Lower Kent Ridge Road,

Singapore 119260)."

BASE COUNT 217 a 88 c 127 g 228 t

ORIGIN

Query Match 60.0%; Score 20.4; DB 14; Length 660;
Best Local Similarity 80.0%; Pred. No. 4.8e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CTCTCTACTTATGATCGGTAAGTGTGAGA 32

DB 413 CTCTCTCCCTTTTGATCGAAGTGTGAGA 384

RESULT 5
BI823138 702 bp mRNA linear EST 04-OCT-2001

LOCUS 603039579F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5180730 5',

mRNA sequence.

BI823138

BI823138.1 GI:15934688

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 702)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

plate: LLAM11450 row: m column: 19

High quality sequence stop: 702.

location/Qualifiers

1. .702

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5180730"

/clone_1lb="NIH_MGC_115"

/lab_host="DH10B"

/note="Organ: pooled brain, lung, testis; Vector:

pcmv-sPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA

source anonymous pool of 6 male brains, age range 23-27; 1

male lung, age 27; and 1 male testis, age 69. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.8 kb,

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C. Gruber (Invitrogen) .. Research Genetics tracking code 021. Note: this is a NIH MCC Library."

BASE COUNT

209 a 155 c 160 g 178 t

Query Match 59.4%; Score 20.2; DB 13; Length 702;
Best Local Similarity 75.8%; Pred. No. 5.9e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACTCTCTACTTATGATCGGCTAGTGAACA 34
|||||
DB 670 ACTGCTGTTATCAATGCTTAATATGACA 702

RESULT 6

LOCUS

DEFINITION BG833308 136 bp mRNA linear EST 23-MAY-2001
951004F05.x2 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,
mRNA sequence.

ACCESSION BG833308 GI:14191078

VERSION
KEYWORDS
SOURCE

EST.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 136)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 8221
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 951004 row: F column: 05.
Location/Qualifiers

FEATURES

source

1. 136
/organism="Zea mays"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/clone_id="951 - BMS tissue from Walbot Lab (GR)"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth
phases"
/lab_host="DH10B"
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI. The
library was prepared by George Rudenko using poly (A)
selected RNA and Universal Riboclone cDNA Synthesis System
(Promega). cDNA was synthesized using both random and
oligo(dT) primers in separate reactions and equipped with
EcoRI adaptors. Library was size-fractionated on agarose
gels (for insert size >400bp) and non-directionally cloned
into EcoRI-digested pUC19 vector. Blue/white selection on
carbenicillin-containing plates was used to recover
positive clones."

BASE COUNT

34 a 35 c 31 g 36 t

Query Match 58.8%; Score 20; DB 12; Length 136;
Best Local Similarity 82.1%; Pred. No. 4.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CCTCTACTTATGATCGGCTAGTGAACA 32
|||||
DB 76 CCTCTCTATGACCGTTTCAGTGAGA 103

RESULT 7
BF542321/c 349 bp mRNA linear EST 11-DEC-2000
LOCUS
DEFINITION UI-R-C2p-se-d-06-0-UI-r1 UI-R-C2p Rattus norvegicus cDNA clone
UI-R-C2p-se-d-06-0-UI 5', mRNA sequence.

ACCESSION BF542321 GI:11633428

VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 349)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA library Preparation: M.B. Soares lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNL (info@image.llnl.gov). IMAGE ID= 1768584
Seq primer: M13 Forward.
Location/Qualifiers

FEATURES

source

1. 349
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-se-d-06-0-UI"
/clone_id="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Bco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996) "

BASE COUNT

121 a 88 c 67 g 73 t

Query Match 58.2%; Score 19.8; DB 12; Length 349;
Best Local Similarity 77.4%; Pred. No. 6.9e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACTCTCTACTTATGATCGGCTAGTGAACA 32
|||||
DB 59 ACTCTCTATTTAAAGAGGTGATATGACA 29

RESULT 8
 AL387696/c
 LOCUS
 DEFINITION AL387696 451 bp mRNA linear EST 03-AUG-2000
 MbBC44B09F1 MbBC Medicago truncatula cDNA clone MbBC44B09 T3, mRNA
 sequence.
 ACCESSION AL387696
 VERSION AL387696.1 GI:9687447
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 451)
 Jounet,E.P., Crepeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
 Niebel,A., Carreau,V., Chagnier,O., Kahn,D., Gnanmazzi-Pearson
 V. and Gamas,P.
 Medicago truncatula ESTs from endomycorrhizal roots
 Unpublished (2000)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Jounet, Laboratoire de
 Biologie Molculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 M-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).
 Location/Qualifiers
 FEATURES
 source
 1.451
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MbBC44B09"
 /clone_1ib="MbBC"
 /tissue_type="arbuscular mycorrhiza"
 /dev_stage="harvested 3 weeks post inoculation with Glomus
 intraradices"
 /note="Vector: pBluescript PSK; Site 1: EcoRI; Site 2:
 XhoI; M. truncatula sterilised seeds were germinated for
 72h at 25 C, before transplanting into a 1/3 Spoilsee soil
 2/3 calcined Terragreen mix in the presence of onion
 root fragments colonized by the arbuscular mycorrhizal
 fungus Glomus intraradices (Schenck & Smith, isolate LpA8
). The plants were watered every day and twice a week with
 a modified nutrient Long Ashton solution without phosphate
 but with a high level of nitrate. After 3 weeks RNA was
 extracted from whole root systems. cDNA was prepared from
 polyA+ enriched RNA. The cDNA was directionally ligated
 into Uni-zap XR vector from Stratagene and packaged using
 Gigapack Gold packaging extracts. Plasmids containing cDNA
 inserts were mass-excised from phage stocks using Exsist
 helper phage and propagated in SOLR cells. Clone ordering
 and sequencing was performed by the Centre National de
 Sequencage (Genoscope, Evry, France). Note : EST may be of
 fungal origin."

BASE COUNT 141 a 77 c 106 g 125 t 2 others
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 9; Length 451;
 Best Local Similarity 77.4%; Pred. No. 7.4e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 137 TGCTCTTATGATGATGCTTGTGAATA 107
 RESULT 9
 BM209528/c

LOCUS BM209528 502 bp mRNA linear EST 31-JAN-2002
 DEFINITION C0649B07-3 NIA Mouse Trophoblast Stem Cell cDNA Library (long) Mus
 musculus cDNA clone C0649B07 3', mRNA sequence.
 ACCESSION BM209528
 VERSION BM209528
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 502)
 Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Luo,A.,
 Tanaka,T., Kunath,T., Rossant,J. and Ko,M.S.H.
 Systematic Analyses of NIA Mouse Trophoblast Stem Cell cDNA Library
 (long)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 Plate: C0649 row: B column: 07
 Seq primer: -21M3 Forward
 High quality sequence stop: 502
 POLYA=Yes.
 FEATURES
 source
 1.502
 /organism="Mus musculus"
 /strain="B6/EGFP transgenic ICR mice"
 /db_xref="nlaEST:C0649B07-3"
 /db_xref="taxon:10090"
 /clone="C0649B07"
 /clone_1ib="NIA Mouse Trophoblast Stem Cell cDNA Library
 (long)"
 /tissue_type="Trophoblast stem cell"
 /dev_stage="3.5-dpc"
 /lab_host="DH10B"
 /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
 NotI; Mouse cDNA project by the Laboratory of Genetics,
 National Institute on Aging (NIA), Intramural Research
 Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
 a long-transcript enriched cDNA library (Ref. Genome Res.
 11: 1553-1558 (2001). [PMID: 11544191]. Total RNAs were
 obtained from Dr. Janet Rossant and Tilo Kunath (Samuel
 Lunenfeld Research Institute, Canada). Double-stranded
 cDNAs were synthesized with an Oligo(dt) primer
 [Invitrogen:
 5'-pGACTGTTCTTAGATCGAGCGGCCCTTTT-3'] from
 4 ug of total RNA, treated with T4 DNA polymerase, and
 purified by ethanol-precipitation. The cDNAs were ligated
 to lone-linker IL-SalI, purified by phenol/chloroform, and
 separated from free linkers by Centricon 100. Then, the
 cDNAs were amplified by long-range high fidelity PCR using
 Ex Taq polymerase (Takara) with a primer SalI-4-S. The
 products were purified by phenol/chloroform and Centricon
 100. The cDNAs were digested with SalI and NotI enzymes
 and cloned into SalI/NotI site of pSPORT1 plasmid vector.
 The DH10B E. coli host was transformed with the ligation
 mixture by the standard chemical method. The average
 insert size is about 2.6 kb. The library was constructed
 by Yulan Piao (NIA)."

BASE COUNT 103 a 136 c 148 g 115 t
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 13; Length 502;
 Best Local Similarity 77.4%; Pred. No. 7.7e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 487 ACCTCTATTATGCTGTGCTGTGAGA 457

<hr/>					
RESULT	10				
LOCUS	BOS59633/c		514 bp	mRNA	linear EST 20-JUN-2002
DEFINITION	H4059D12-3 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone				
ACCESSION	H4059D12.3				mRNA sequence.
VERSION	BOS59633				
KEYWORDS	BOS59633.1 GI:21460518				
SOURCE	EST.				
ORGANISM	Mus musculus				house mouse.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 514)				
JOURNAL	Vanduren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin,				
COMMENT	P.R., Stegg,C.A., Basse,y,U., Alba,K., Hametant,T., Kargul,G.J., Luo,A.G. and Ko,M.S.H. Assembly, verification, and initial annotation of NIA 7.4K mouse cDNA clone set Unpublished (2002) Contact: Yong Qian Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA Email: cdna@igsn.grc.nia.nih.gov This clone set has been freely distributed to the community. Please visit http://igsn.grc.nia.nih.gov/cDNA_NIA_7.4K.html for details. Plate: H4059 row: D column: 12 Seq primer: -21M13 Forward High quality sequence stop: 514 POLYA=yes. Location/Qualifiers 1..514 /organism="Mus musculus" /strain="C57BL/6" /db_xref="taxon:H4059D12-3" /db_xref="taxon:10090" /clone="H4059D12" /clone_1lb="NIA Mouse 7.4K cDNA Clone Set" /bex="mixed" /dev_stage="mixed" /lab_host="DH10B" /note="Vector: pSPORF1; Site 1: SalI; Site 2: NotI. This clone is among a rearrayed set of 7,407 clones from more than 20 cDNA libraries."				
BASE COUNT	106 a 142 C 148 g 118 t				
ORIGIN					
Query Match		58.2%	Score 19.8;	DB 14;	Length 514;
Best Local Similarity	77.4%;	Pred. No. 7.7e+02;			
Matches 24; Conservative	0;	Mismatch	7;	Indels	0;
Gaps	0;				
DQ	2 ACTGCTACTATGCATCGTGTCTGGAGA 32				
Db	487 ACCTCTATTATTCGTCTGTGTAGGA 457				
RESULT 11					
LOCUS	BB764031		524 bp	mRNA	linear EST 17-OCT-2001
DEFINITION	BB764031 RIKEN full-length enriched, Bl6 F10Y cells Mus musculus				
ACCESSION	CMDA clone G370032Ff1.3'				mRNA sequence.
VERSION	BB764031				
KEYWORDS	BB764031.1 GI:16208434				
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 524)				
JOURNAL	Akimura,T., Atrakawa,T., Carinici,P., Furuno,M., Hanagaki,T.,				
COMMENT	Kojima,Y., Hiramoto,K., Hiroaka,T., Hitozane,T., Imotani,K., Ishii, Y., Ito,M., Kawai,J., Komatsu,H., Koude,M., Matsuyama,T., Teshima, Takamizawa,N., Tanabe,T., Takahashi,K., Taniguchi,K., Tanisaka,E., Taniguchi,K., Tomita,K., Ueda,K., Watanabe,K., Yokoyama,C., Yoda,S., Yoshida,K., Yoshimizu,K., Yamada,K., Nakamura,K., Okazaki,Y., Oikido,T.,				

TITLE	Unpublished (2001)			
JOURNAL	Contact: Yoshihide Hayashizaki			
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, url:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujisawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuhiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.			
FEATURES	Location/Qualifiers			
SOURCE	1. 524 /organism="Mus musculus" /dbxref="C57BL/6J" /dbxref="taxon:10090" /clone="G370032F11" /clone_1lb="RIKEN full-length enriched, B16 F10Y cells" /cell_type="B16 F10Y cells" /note="pooled tissues; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mxed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mxed), (tissue_type=hippocampus, dev_stage=adult, sex=mxed), (tissue_type=whole body, dev_stage=3 days embryo, sex=mxed), (tissue_type=lung, dev_stage=13 days embryo, sex=mxed)"			
BASE COUNT	120 a	152 c	147 g	105 t
ORIGIN				
Query Match	58.2%	Score 19.8;	DB 10;	Length 524;
Best Local Similarity	77.4%;	Pred. No. 7.8e+02;		
Matches 24;	Conservative	0;	Mismatches 7;	Indels 0;
Gaps 0;				
Db	34	ACCTCTATTATTCGTCTGTCTGTGAGA	64	
0y	2	ACTGCTACTTATCGATCGGTACGTGAGA	32	
1y	1			
2y	1			
3y	1			
4y	1			
5y	1			
6y	1			
7y	1			
8y	1			
9y	1			
10y	1			
11y	1			
12y	1			
13y	1			
14y	1			
15y	1			
16y	1			
17y	1			
18y	1			
19y	1			
20y	1			
21y	1			
22y	1			
23y	1			
24y	1			
25y	1			
26y	1			
27y	1			
28y	1			
29y	1			
30y	1			
31y	1			
32y	1			
33y	1			
34y	1			
35y	1			
36y	1			
37y	1			
38y	1			
39y	1			
40y	1			
41y	1			
42y	1			
43y	1			
44y	1			
45y	1			
46y	1			
47y	1			
48y	1			
49y	1			
50y	1			
51y	1			
52y	1			
53y	1			
54y	1			
55y	1			
56y	1			
57y	1			
58y	1			
59y	1			
60y	1			
61y	1			
62y	1			
63y	1			
64y	1			
65y	1			
66y	1			
67y	1			
68y	1			
69y	1			
70y	1			
71y	1			
72y	1			
73y	1			
74y	1			
75y	1			
76y	1			
77y	1			
78y	1			
79y	1			
80y	1			
81y	1			
82y	1			

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 530)
 Akimura,T., Arakawa,T., Carinci,P., Furuno,M., Hangaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL
COMMENT

Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Osawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
 *Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
 e mouse tissues.

FEATURES
source

Location/Qualifiers
 1. 530
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G270012B20"
 /clone_1ib="RIKEN full-length enriched, melanocyte"
 /cell_type="melanocyte"
 /note="pooled tissues; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"
 BASE COUNT 121 a 157 c 137 g 115 t
 ORIGIN

Query Match 58.2%; Score 19.8; DB 10; Length 530;
 Best Local Similarity 77.4%; Pred. No. 7.8e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 13
 AZ331947/c 596 bp DNA linear GSS 29-SEP-2000
 LOCUS

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

IM0060J06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0060J06 F, DNA sequence.
 AZ331947
 AZ331947.1 GI:10395130
 GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 596)

JOURNAL
COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0060 row: J column: 06
 Seq primer: CCGTGTAAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 596.
 Location/Qualifiers
 1. 596
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0060J06"
 /clone_1ib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (GI:4732114|gb|AF12072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

FEATURES
source

Location/Qualifiers
 1. 596
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0060J06"
 /clone_1ib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (GI:4732114|gb|AF12072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 58.2%; Score 19.8; DB 17; Length 596;
 Best Local Similarity 77.4%; Pred. No. 8.1e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

RESULT 14
 AZ831669 611 bp DNA linear GSS 20-FEB-2001
 LOCUS

DEFINITION 2M011H02R Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 ACCESSION clone UNGC2M011H02 R, DNA sequence.
 VERSION A2831669
 KEYWORDS A2831669.1 GI:13001577
 SOURCE GSS.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 611)
 REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,
 and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0111 row: H column: 02
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 611.
 Location/Qualifiers
 1. 611
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNG2M011H02"
 /clone_1db="Mouse 10kb plasmid UNGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid RI. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 153 a 123 c 182 g 153 t
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 17; Length 611;
 Best Local Similarity 77.4%; Pred. No. 8.2e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 4 TCCTCTACTATTCGATCGTACTGTGAGACA 34
 Db 252 TCCTCTGATTCGATCGTACTGTGAGACA 282
 RESULT 15
 BM945933 615 bp mRNA linear EST 14-MAR-2002
 L6CUS

DEFINITION UI-M-EMO-bv1-f-05-0-UI.r1 NIH BMAP_EMO Mus musculus cDNA clone
 IMAGE:5692804 5', mRNA sequence.
 ACCESSION BM945933
 VERSION BM945933.1 GI:19429518
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 615)
 REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@emil.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1. 615
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5692804"
 /clone_1db="NIH BMAP EMO"
 /tissue_type="whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (TI phage resistant)"
 /note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according to
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CAGCAGCAGC. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous
 System', supported by National Institute of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 122 a 178 c 175 g 140 t
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 14; Length 615;
 Best Local Similarity 77.4%; Pred. No. 8.2e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 2 ACTCTCTACTATTCGATCGTACTGTGAGACA 32
 Db 527 ACTCTCTACTATTCGATCGTACTGTGAGACA 557
 Search completed: May 11, 2003, 04:54:25
 Job time : 550.556 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic & nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 349.519 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33
Sequence: 1 gctgtgtctactctctctctctctcaccagca 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_hcg: 3: gb_in: 4: gb_ov: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_stc: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_or: 21: em_ov: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_stc: 28: em_un: 29: em_vl: 30: em_hcg_hum: 31: em_hcg_inv: 32: em_hcg_other: 33: em_hcg_mus: 34: em_hcg_pln: 35: em_hcg_rtd: 36: em_hcg_mam: 37: em_hcg_vrt: 38: em_sy: 39: em_hggo_hum: 40: em_hggo_mus: 41: em_hggo_other:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	33	100.0	33	6	AX036750	Sequence
2	24	72.7	180577	2	AC115145	Rattus no
3	23.4	70.9	31069	2	AC130253	Rattus no
4	23.4	70.9	147295	3	AGA439060	Alu33060 Anopheles
5	23.4	70.9	149788	8	AT76621	Alu58883 Arabidops
6	23.4	70.9	173535	5	AC021316	AC021316 Homo sapi
7	23.4	70.9	194974	2	AC107787	AC107787 Mus muscu
8	23.4	70.9	336062	2	AC091229	AC091229 Rattus no
9	23.4	70.9	345243	2	AC091347	AC091347 Rattus no
10	23.2	70.3	177899	10	AC123042	AC123042 Mus muscu
11	23	69.7	60921	9	AL391706	AL391706 Human DNA
12	23	69.7	224860	2	AC124546	AC124546 Mus muscu
13	23	69.7	229158	2	AC124707	AC124707 Mus muscu
14	22.6	68.5	234900	10	AL731717	AL731717 Mouse DNA
15	22.4	67.9	664	9	HS434005	AJ340055 Homo sapi
16	22.4	67.9	48042	2	AC020217	AC020217 Drosophila
17	22.4	67.9	83656	9	AC007401	AC007401 Homo sapi
18	22.4	67.9	103237	2	AC110484	AC110484 Rattus no
19	22.4	67.9	150765	2	AC094766	AC094766 Rattus no
20	22.4	67.9	167002	2	AC073041	AC073041 Homo sapi
21	22.4	67.9	169705	3	AC104704	AC104704 Drosophila
22	22.4	67.9	171907	2	AC116538	AC116538 Drosophila
23	22.4	67.9	181656	2	AC124900	AC124900 Rattus no
24	22.4	67.9	184620	2	AC096290	AC096290 Rattus no
25	22.4	67.9	197944	2	AC112748	AC112748 Rattus no
26	22.4	67.9	236303	9	AF168681	AF168681 Homo sapi
27	22.4	67.9	279679	3	AE003527	AE003527 Drosophila
28	22.2	67.3	1800	8	YSCMAT1A	L00060 Yeast (S. ce
29	22.2	67.3	2430	8	YSCMAT2A	V01313 Yeast Gene
30	22.2	67.3	2520	8	YSCMAT3A	L00059 Yeast (S. ce
31	22.2	67.3	2563	8	YSCMAT4A	X63853 S. cerevisia
32	22.2	67.3	9818	8	SCMATLOC	248613 S. cerevisia
33	22.2	67.3	32481	8	SC8270	AL160262 Human DNA
34	22.2	67.3	53596	9	AL160262	Continuation (4 of
35	22.2	67.3	71804	2	AC099209_3	AC067861 Homo sapi
36	22.2	67.3	180978	2	AC067861	Continuation (3 of
37	22.2	67.3	190399	2	AC127960	AC127960 Rattus no
38	22.2	67.3	316613	8	SCCHR11	X59720 S. cerevisia
39	22.2	66.7	1424	8	ASNRNVA	D28341 Aspergillus
40	22	66.7	70072	2	AC109210	AC109210 Mus muscu
41	22	66.7	100810	2	AC011991	AC011991 Homo sapi
42	22	66.7	110000	2	AC129176_2	Continuation (3 of
43	22	66.7	150980	2	AC107206	AC107206 Oryza sat
44	22	66.7	157250	2	AC027264	AC027264 Homo sapi
45	22	66.7	158770	9	AC007321	AC007321 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX036750 33 bp DNA
DEFINITION Sequence 16 from Patent WO0058485.
ACCESSION AX036750
VERSION AX036750.1 GI:11226259
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE
1 (bases 1 to 33)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimera expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 16 05-OCT-2000;

Pred. No. is the number of results predicted by chance to have a

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
 ; GRUBER VERONIQUE (FR)
 FEATURES
 Location/Qualifiers
 1..33
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Guide deoxynucleotide building block G2"
 BASE COUNT
 4 a 9 c 4 g 16 t
 ORIGIN
 Query Match 100.0%; Score 33; DB 6; Length 33;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 GCTGTGTTACTTTCTCTCTATTCAGCCA 33
 1 GCTGTGTTACTTTCTCTCTATTCAGCCA 33
 RESULT 2
 AC115145/c
 LOCUS
 DEFINITION
 AC115145 180577 bp DNA linear HTG 17-JUL-2002
 Rattus norvegicus clone CH230-201L12, *** SEQUENCING IN PROGRESS
 *** 53 unordered pieces.
 AC115145
 AC115145.2 GI:21743797
 HTG: HTGS PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 180577)
 1 Mzuy,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alebrooke,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabiela,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Eten,C., Liu,J., Liu,W., Loulesged,H.,
 Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,B., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newson,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokemko,S., Oguh,M., Okwundu,G.,
 Otagunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tasey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Woodson,C., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Direct Submission
 2 (bases 1 to 180577)
 Morley,K.C.
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 180577)
 Morley,K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 13, 2002 this sequence version replaced gi:19482203.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GDB
 Center clone name: CH230-201L12
 ----- Summary Statistics -----
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 132770 bases at least Q40
 Consensus quality: 139349 bases at least Q30
 Consensus quality: 14434 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 53 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 1228: contig of 1228 bp in length
 1229 1338: gap of unknown length
 1329 2518: contig of 1199 bp in length
 2519 2618: gap of unknown length
 2619 3651: contig of 1033 bp in length
 3652 3751: gap of unknown length
 3752 5280: contig of 1533 bp in length
 5291 5390: gap of unknown length
 5391 6657: contig of 1267 bp in length
 6658 6757: gap of unknown length
 6758 7848: contig of 1091 bp in length
 7849 7949: gap of unknown length
 7949 9351: contig of 1403 bp in length
 9352 9451: gap of unknown length
 9452 10649: contig of 1198 bp in length
 10650 10749: gap of unknown length
 10750 12018: contig of 1269 bp in length
 12019 12118: gap of unknown length
 12119 13473: contig of 1355 bp in length
 13474 13573: gap of unknown length
 13574 15170: contig of 1597 bp in length
 15171 15270: gap of unknown length
 15271 16299: contig of 1029 bp in length
 16300 16399: gap of unknown length
 16400 17636: contig of 1237 bp in length
 17637 17737: gap of unknown length
 17737 19787: contig of 2051 bp in length
 19788 19887: gap of unknown length
 19888 21207: contig of 1320 bp in length
 21208 21308: gap of unknown length
 21308 22814: contig of 1507 bp in length
 22815 22914: gap of unknown length

22915	24265	contig of 1349 bp in length
24264	24365	gap of unknown length
24364	25847	contig of 1484 bp in length
25847	25947	gap of unknown length
25948	27776	contig of 1829 bp in length
27777	27876	gap of unknown length
27877	30627	contig of 2751 bp in length
30628	30727	gap of unknown length
30728	32028	contig of 1302 bp in length
32030	32128	gap of unknown length
32130	33789	contig of 1660 bp in length
33790	33889	gap of unknown length
33890	35385	contig of 1496 bp in length
35386	35485	gap of unknown length
35486	37118	contig of 1634 bp in length
37120	37218	gap of unknown length
37220	38423	contig of 1204 bp in length
38424	38523	gap of unknown length
38524	39985	contig of 1462 bp in length
39986	40085	gap of unknown length
40086	41980	contig of 1895 bp in length
41981	42080	gap of unknown length
42081	45255	contig of 3173 bp in length
45254	45355	gap of unknown length
45355	48312	contig of 2959 bp in length
48313	48412	gap of unknown length
48413	51843	contig of 3431 bp in length
51844	51943	gap of unknown length
51944	55120	contig of 3177 bp in length
55121	55220	gap of unknown length
55221	59133	contig of 3913 bp in length
59134	59233	gap of unknown length
59234	61562	contig of 2329 bp in length
61563	61662	gap of unknown length
61663	64287	contig of 2625 bp in length
64288	64387	gap of unknown length
64388	66172	contig of 1785 bp in length
66173	66272	gap of unknown length
66273	69183	contig of 2911 bp in length
69184	69283	gap of unknown length
69284	73838	contig of 4555 bp in length
73839	73938	gap of unknown length
73939	77232	contig of 3794 bp in length
77233	77832	gap of unknown length
77833	80598	contig of 2766 bp in length
80599	80698	gap of unknown length
80699	84519	contig of 3821 bp in length
84520	84619	gap of unknown length
84620	88147	contig of 3528 bp in length
88148	88247	gap of unknown length
88248	92434	contig of 4187 bp in length
92435	92534	gap of unknown length
92535	98849	contig of 6315 bp in length
98850	98949	gap of unknown length
98950	104696	contig of 5747 bp in length
104697	104796	gap of unknown length
104797	110912	contig of 6116 bp in length
110913	111013	gap of unknown length
111013	118020	contig of 7008 bp in length
118021	118120	gap of unknown length
118121	124423	contig of 6309 bp in length
124430	124529	gap of unknown length
124530	130536	contig of 6007 bp in length
130537	130636	gap of unknown length
130637	133183	contig of 8547 bp in length
133184	139283	gap of unknown length
139284	147688	contig of 8402 bp in length
147689	147785	gap of unknown length
147786	156520	contig of 8735 bp in length
156521	156620	gap of unknown length
156621	166272	contig of 9553 bp in length
166273	166277	gap of unknown length
166278	180577	contig of 14304 bp in length

FEATURES	Location/Qualifiers
SOURCE	1. 180577
Query Match	72.7%; Score 24; DB 2; Length 180577;
Best Local Similarity	84.4%; Pred. No. 64;
Matches	27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy	2 CTTGTTACTTTCTCTCTATTCAGGCCA 33
Db	67869 CTTGTTACTTTCTCTCTATTCAGGCCA 67838
AC130253	
LOCUS	
DEFINITION	AC130253 31069 bp DNA linear HTG 09-AUG-2002
ACCESSION	AC130253
VERSION	AC130253.1 GI:22164860
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 31069)
	Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralung,H.C., Are,J.R., Ayala,M., Banks,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Briveau,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delany,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Goriell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsed,H., Lotado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M., Mier,G., Minter,Z., Mitchell,T., Mohabbat,K., Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenko,S., Ogun,M., Okunnu,G., Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Qules,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruliz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Szatek,A., Taber,P., Tamerias,A., Tamezis,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,A., Thomas,N., Thomas,S., Umanai,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wlezyk,R., Woden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinscock,G. and Gibbs,R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCES	2 (bases 1 to 31069)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (09-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

EHAGSITGDKNLPAGTKILSYAEKASTNLVEKFTQSVRIIVQDVKEGSGPGMSRE
ELLETERHEAARLLEESYDTAKALVINMANKYSGSKSRNIFNNRYPMKIMIKXVI
RLQETWTLVEIPEKLETVIPAFVACSIMKSGSGCVTSAKLAEEAERER
MERLEAMTTOIEONTOMINDYRLKKYTGARNIRELSEYGNISKYPIPRRTM
LKMIDIMEDPRMEVCHVDN"
640..654
/gene="4F11.9"
/number=1
/gene=".1575
/gene="4F11.9"
/number=1
1576..1893
/gene="4F11.9"
/number=2
1894..2258
/gene="4F11.9"
/number=2
2259..3527
/gene="4F11.9"
/number=3
3528..4486
/gene="4F11.9"
/number=3
4487..4564
/gene="4F11.9"
/number=4
4565..4638
/gene="4F11.9"
/number=4
4639..4821
/gene="4F11.9"
/number=5
4822..4889
/gene="4F11.9"
/number=5
4890..5087
/gene="4F11.9"
/number=6
5088..5159
/gene="4F11.9"
/number=6
5160..5267
/gene="4F11.9"
/number=7
5268..5342
/gene="4F11.9"
/number=7
5343..5465
/gene="4F11.9"
/number=8
18352..21161
/gene="4F11.10"
join(18352..18654,21057..21161)
/gene="4F11.10"
join(18352..18654,21057..21161)
/gene="4F11.10"
/codon_start=1
/product="putative cytoskeletal regulator"
/protein_id="CAD27753.1"
/db_xref="GI:19577371"
/translation="MPFVKRVVTPKYIARSSKILQPYAAPTIPVADNEFALTNT
LSNALROLASLVYISNOIFTELHQLSESVRSIGIKORIDRLSAKVDQDDPQVTV
SSISRSRSTVSNVNLITCPSMAHAPGCRHT"
18352..18654
/gene="4F11.10"
/number=1
18655..21056
/gene="4F11.10"
/number=1
21057..21161
/gene="4F11.10"
/number=2
23341..31440
gene

mRNA
CDS
/gene="11N17.1.1"
join(23341..23391,23609..23881,23940..23984,25261..25608,
25712..25945,26387..28585,28647..29090,30097..31440)
/gene="11N17.1.1"
join(23341..23391,23609..23881,23940..23984,25261..25608,
25712..25945,26387..28585,28647..29090,30097..31440)
/gene="11N17.1.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD27754.1"
/db_xref="GI:19577372"
/translation="MSQFQRLCEKLAHIGSDLDTFVQIKHYTTKHVDTGLFTVA
TSEVQELVEAARVPAVIAEMDRIMINIGNSTEAPICPVLSROKATAPITST
WTSKYLCEELFLFARSSSLVTIPTHSHQPAVODLRMTTEALADVTGACSAK
IASHSINNSVLPYTESPTDLDAPTNSMAAGDETDLDAITLAEIGIPSSNIG
DRAVDHLPSPAOCRIILASKPAETIKIDTSGRAPRMSSTKSLVCTYTTOKEGD
MAKPEBDYRRRSRLRRPRGRKRNITASSDQRIEAVINKGPESTVEHYPCDGE
LMPAIRSKSGDILRKEPSTSAEMASFNMSHPNSLQMGKRLRMNRDSSASORD
KQODIDFNIHDTTLAROPSVGSSRRRGGSTTDREKRLSHDRKPSYSSERSSTGILG
GTAAYLPASINVEKLESTIRORRTALGNRDEPHSSGNMSASSESGRTIGSEIT
TNTHPKSSASSTLNSHNPISSSAPPSIVRRRRPNTSASSVTSEGITPDLQTF
DYHDEGGETSSVSCDTEGCVTSPFVHDSGLKTLKCEEPMTPLOSTTALSITSFSSG
NTTVSDYVYKSGSTSTTSSAGTICITLADGDSVSAASNLKPIPRKSLTYLN
KSNSTASNGTLERSYSSSTLSTGTIKNGVLIQKEVILIKIQENNNLIDVADG
NAPKVAAGTGAVOSSSEAVSSKLPPTLAPESBESDVECVARTELKYKTTINTSRI
PMCLITPTNSDDEQPPAGSATKAKATORDNSTKAGIGACGRAPPTTESTSEILK
PDPRKHSLLPLNNVDRIKALPHLKSPTEBDATGSGSTTTSKDEBEAAGGGE
VYI8KGRTEPTELTVRRLATVLSGNINTEFEVYSINLPCNICEGANLIDBCK
GVYBESPAATELSBDALQSGDGGSGPGLGGIGROALKGARVLDANGRAIF
SSDSLRRRGAHTTFAFGPCQQQ000QVNGSAGGTSTNVPITGAPRLPTTRTLL
LKRPLVASRYGTPRIGPAPVAPAKTIVATILPNSAKPAQOPPLRPDARELASRS
FSPYGADVSRGDRGGAAYAGAAPAAVQARSVAGTAAVAGHAKLYPEGASHLD
RSIRNWVPDKLTVOGNPSAATTLPRPYARTVYAGADYAPAPLASAGGASFP
ASPLVDDTAPDPDHSIXAIPNKYKESPLAGAAPKMPMSDLSIILAEOSELPIPKCO
TNPRTSTSKEDBALGALGPLHRLLSNVSLGNSPVHSGRSTRELBESSQPAAG
GTPRGHSHVANSVEVPNTCSORLGTPTKSLMDPKKLLAHGTKTHASPGSKMSAVM
LKKSKESAATRPKQDPGGSPASSGMAILDMASAPKMYFRIIAQGGYGSPTGA
GGGCGGCGGGAAGRSNMYNNMRTGVISTALPEANSEEDIYPPAPATATTSVEREE
PVASGKAPESSESGGGVVEETISFKENIFLKEDENNFMGEVPRMTKSPAGFSTV
DGRKDMAGGRLLQARDOGADEAASKQALETAL"
23341..23391
/gene="11N17.1.1"
/number=1
23392..23608
/gene="11N17.1.1"
/number=1
23609..23881
/gene="11N17.1.1"
/number=2
23882..23939
/gene="11N17.1.1"
/number=2
23940..23984
/gene="11N17.1.1"
/number=3
23985..25260
/gene="11N17.1.1"
/number=3
25261..25608
/gene="11N17.1.1"
/number=4
25609..25711
/gene="11N17.1.1"
/number=4
25712..25945
/gene="11N17.1.1"
Query Match 70.9%; Score 23.4; DB 3; Length 147295;
Best Local Similarity 81.8%; Pred No. 1.1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTGTTACTTTCTTCTTCTATTCACGCCA 33
|||||

Db 97476 GCTGTGCTGCTTGCATTTCATTCACGCCA 97444

RESULT 5
ATT6G21/c 149788 bp DNA linear PLN 14-MAR-2001
LOCUS Arabidopsis thaliana DNA chromosome 5, BAC clone T6G21 (ESSA
DEFINITION project).
ACCESSION AL589883
VERSION AL589883.1 GI:13374848
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 149788)
Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W.,
Rudd, S., Lemcke, K. and Mayer, K.F.X.
Unpublished
2 (bases 1 to 149788)
EU Arabidopsis sequencing project.
DIRECT SUBMISSION
Submitted (12-MAR-2001) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
Mayer@mips.gsf.de Project Coordinator: Mike Bevan, Molecular
Genetics Department, Cambridge Laboratory, John Innes Centre,
Colney Lane, NR4 7UJ Norwich UK, E-mail: michael.bevan@brc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
Location/Qualifiers
1. 149788
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
5957..8056
/gene="At5g21900"
join(5957..6637,6716..7152,7250..7411,7501..7576,
7657..7812,7934..8056)
/gene="At5g21900"
/note="similarity to DNA excision repair protein,
Schizosaccharomyces pombe, PIR:T41312
contains Myb DNA-binding domain repeat signatures
AA502-510"
/codon_start=1
/product="DNA excision repair protein"
/protein_id="CAC34483.1"
/db_xref="GI:13374849"
/translation="MNSGDPNDRENRRIINTDEAIDALFMATEBAVTRDPLSGPON
ODVMEERKSTADVTKGSDYMDYDEKRNFRKAKPALELSDLKIVKTEBEVE
EVIKRDPEKSRKAKESIKKROKISTSDMEIDLNRRKAPSLVEISAVLAKNPLA
IKSLKLVFDHRLKRLSTVSGLSGFDRLMLBLLEDPSSEICANCVQVEDDVKIF
SCDQSVLKVLIIDLGRSMYDTYINQFRKAPGFSPLTTLISQAFCLTDNALLI
SKSPPLQVINTLTCESLLTYRALIILDKFGSTLIGSGCOGIKHGFSSSLYF
EKLVYLSVAGVSVNDGVRSFWFRFSILTDLSLANCVTEDECMHIGRYCKLEA
LDITDLKLTDKSLLEFTEGCRVYKSLKLTNSRSDSCIAFLVSGGSIAREICLVY
RDVGPETAFSLAKYCKMLQFLDLSMCKRLKEDDLRLRCCSSLSQSLKLGMYQVEET
YLERLSRDVHITGLKLTSLYHLNDPSPVGAKEP"
5957..6637
/gene="At5g21900"
/number=1
6638..6715
/gene="At5g21900"
/number=1
6716..7152
/gene="At5g21900"
/number=2
7153..7249
/gene="At5g21900"
/number=2
7250..7411

introns
/gene="At5g21900"
/number=3
7412..7500
/gene="At5g21900"
/number=3
7501..7576
/gene="At5g21900"
/number=4
7577..7656
/gene="At5g21900"
/number=4
7657..7812
/gene="At5g21900"
/number=5
7813..7933
/gene="At5g21900"
/number=5
7934..8056
/gene="At5g21900"
/number=6
8057..8845
/gene="At5g21910"
8534..8845
/gene="At5g21910"
/codon_start=1
/product="putative protein"
/protein_id="CAC34484.1"
/db_xref="GI:13374850"
/translation="MGRFTTSKRLAISLVANLSPYITDLSRDEDEEELPSFM
FSPSLTLILLIALLSTVSDPSLTFPDYWIHRLCGSFGLVITILLVFLVLCAS
G"
8534..8845
/gene="At5g21910"
/number=1
9068..10483
/gene="At5g21920"
complement(join(9068..9314,9598..9800,10178..10483))
/gene="At5g21920"
complement(join(9068..9314,9598..9800,10178..10483))
/gene="At5g21920"
/note="similarity to various predicted proteins,
Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAC34485.1"
/db_xref="GI:13374851"
/translation="MEASANEPMKSLKSNPSPGPIPNFVSLSSATQTPVRSNKN
LILLPVAVSVKLIIDFQSLISATEKFGFLHSLSKNPLQEAVALSSBPHLCDE
IRLRNTRVRPAMSNHNGFAVLPGDSVAGLVANGLINLNTVITLVRLVLPSPS
APPALVNPSTLCDDPYLNTFRGFIPPLGGLDLSPIIAFLVNAFTSSAMALPCELPSA
DGVSPASSETKWRRRRLSSHKDRHSPSSASMT"
complement(9068..9314)
/gene="At5g21920"
/number=1
9315..9597
/gene="At5g21920"
/number=1
complement(9598..9800)
/gene="At5g21920"
/number=2
complement(9801..10177)
/gene="At5g21920"
/number=2
complement(10178..10483)
/gene="At5g21920"
/number=3
10484..16439
/gene="At5g21930"
join(10847..10907,10989..11468,11567..11644,12017..12114,
13201..13254,13275..12461,12550..12714,13083..13193,
13335..13403,13695..13804,13981..14086,14288..14354,
14769..14848,15000..15629,15804..15935,16067..16217,
16348..16439)

TITLE
JOURNAL
COMMENT

Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,
Teafaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, I., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 330 Charles Street, Cambridge, MA 02141, USA
On Aug 23, 2002 this sequence version replaced gi:21307379.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

FEATURES

Source

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L4301
Center clone name: 11_J_16

Location/Qualifiers

1. 173535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-11J16"
/clone_1db="RRC1-11 Human Male BAC"
complement (1. 1513)
/rpt_family="L1"
repeat_region
1514. 1823
/rpt_family="AluY"
1824. 1848
/rpt_family="T(A)n"
complement (1849. 2309)
/rpt_family="L1"
repeat_region
complement (2304. 3709)
/rpt_family="L1PBA"
3767. 3821
/rpt_family="TIGR5"
complement (4182. 4475)
/rpt_family="AluSC"
4497. 4517
/rpt_family="AT_rich"
complement (4523. 4779)
/rpt_family="TIGR3 (Golem)"
complement (4777. 4866)
/rpt_family="TIGR3 (Golem)"
complement (5350. 5474)
/rpt_family="L1ME"
complement (7827. 14234)
/rpt_family="L1PA7"
14502. 14555
/rpt_family="L2"
14893. 15214
/rpt_family="L1ME3"
15215. 15519
/rpt_family="Aluub"
15520. 16234
/rpt_family="L1ME3"
16247. 16848
/rpt_family="MLT1E2"
16863. 17022
/rpt_family="L1ME3"
17006. 17121
/rpt_family="L1MA4"
17239. 17445
/rpt_family="L2"
18801. 19118
/rpt_family="L1T855"
complement (19119. 19164)
/rpt_family="MLT2D"
19177. 19462

/rpt_family="TIGR3b"
19463. 19537
/rpt_family="T(A)n"
complement (19538. 19841)
/rpt_family="AluY"
19842. 20687
/rpt_family="TIGR3b"
complement (20765. 21093)
/rpt_family="MLT2D"
21977. 22064
/rpt_family="CATARA)n"
complement (22066. 22279)
/rpt_family="TIGR3 (Golem)"
complement (22186. 22190)
/note="<30 qual SNGL region"
complement (22262. 22271)
/note="<30 qual SNGL region"
complement (22279. 22409)
/rpt_family="TIGR3 (Golem)"
22844. 23153
/rpt_family="Aluub"
23234. 23302
/rpt_family="CATARA)n"
23334. 23497
/rpt_family="MLT1J2"
complement (23589. 23593)
/note="<30 qual SNGL region"
complement (23619. 23623)
/note="<30 qual SNGL region"
23639. 23815
/rpt_family="MLT1J1"
complement (24116. 24458)
/rpt_family="L1ME3"
complement (24459. 24782)
/rpt_family="L2"
complement (24731. 24735)
/note="<30 qual SNGL region"
complement (24742. 24748)
/note="<30 qual SNGL region"
complement (24767. 24772)
/note="<30 qual SNGL region"
complement (24781. 24789)
/note="<30 qual SNGL region"
complement (24783. 25086)
/rpt_family="AluSg"
complement (24802. 24812)
/note="<30 qual SNGL region"
complement (24818. 24838)
/note="<30 qual SNGL region"
complement (24862. 24870)
/note="<30 qual SNGL region"
complement (25087. 25675)

Query Match 70.9%; Score 23.4; DB 9; Length 173535;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTTTACTTTCTTCTTATTCACCA 33
DB 80832 GCTGCTTAGACATTCTCTTATTCACCA 80800

RESULT 7
AC107787 194974 bp DNA linear HTG 12-APR-2002
LOCUS Mus musculus clone RP23-432P10, WORKING DRAFT SEQUENCE, 34 ordered
pieces.
ACCESSION AC107787
VERSION AC107787.2 GI:20143689
KEYWORDS HTG, HTGS_PHASE2, HTGS_DRAFT, HTGS_FULLTOP.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194974)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-432710
Unpublished

2 (bases 1 to 194974)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Lander, E., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Menus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rettig, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194974)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Lander, E., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Menus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Rettig, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (12-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 12, 2002 this sequence version replaced gi:18308489.
All repeats were identified using RepeatMasker:
Smith, A. P. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: L20471
Center clone name: 432 P 10
Summary Statistics
Sequencing vector: Plasmid, n/a, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.960731

Consensus quality: 185565 bases at least Q40
Consensus quality: 188852 bases at least Q30
Consensus quality: 190073 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 191674; sum-of-contigs
Quality coverage: 8.9 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 34 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.

This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preceeded.

1 1022 1121: contig of 1021 bp in length
1022 1121: gap of 100 bp
1122 1606: contig of 485 bp in length
1607 1706: gap of 100 bp
1707 2355: contig of 649 bp in length
2356 2455: gap of 100 bp
2456 3109: contig of 654 bp in length
3110 3209: gap of 100 bp
3210 4151: contig of 942 bp in length
4152 4251: gap of 100 bp
4252 4930: contig of 679 bp in length
4931 5030: gap of 100 bp
5031 6133: contig of 1093 bp in length
6124 6223: gap of 100 bp
6224 6989: contig of 766 bp in length
6990 7089: gap of 100 bp
7090 7786: contig of 707 bp in length
7787 7896: gap of 100 bp
7897 9525: contig of 1629 bp in length
9526 9625: gap of 100 bp
9626 10896: contig of 1271 bp in length
10897 10986: gap of 100 bp
10987 13026: contig of 2030 bp in length
13027 13126: gap of 100 bp
13127 45197: contig of 32071 bp in length
45198 45297: gap of 100 bp
45298 47519: contig of 2222 bp in length
47520 47619: gap of 100 bp
47620 49698: contig of 2079 bp in length
49699 49798: gap of 100 bp
49799 51950: contig of 2152 bp in length
51951 52050: gap of 100 bp
52051 54612: contig of 2562 bp in length
54613 54712: gap of 100 bp
54713 57085: contig of 2373 bp in length
57086 57185: gap of 100 bp
57186 60203: contig of 3018 bp in length
60204 60303: gap of 100 bp
60304 63082: contig of 2779 bp in length
63083 63182: gap of 100 bp
63183 66877: contig of 3695 bp in length
66878 66977: gap of 100 bp
66978 71786: contig of 4809 bp in length
71787 71886: gap of 100 bp
71887 77847: contig of 5961 bp in length
77848 77947: gap of 100 bp
77948 82080: contig of 4133 bp in length
82081 82180: gap of 100 bp
82181 88846: contig of 6666 bp in length
88847 88946: gap of 100 bp
88947 94785: contig of 5839 bp in length
94786 94885: gap of 100 bp
94886 102269: contig of 7384 bp in length
102270 102369: gap of 100 bp
102370 110158: contig of 7789 bp in length
110159 110258: gap of 100 bp

```

* 110259 120287: contig of 10029 bp in length
* 120288 120387: gap of 100 bp
* 120388 134724: contig of 14337 bp in length
* 134725 134824: gap of 100 bp
* 134825 151821: contig of 16997 bp in length
* 151822 151921: gap of 100 bp
* 151922 174022: contig of 22101 bp in length
* 174023 174122: gap of 100 bp
* 174123 193613: contig of 19490 bp in length
* 193613 193713: gap of 100 bp
* 193713 194974: contig of 1262 bp in length.
Location/Qualifiers
1. 194974
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-432P10"
/clone_1db="RP21-23 Female Mouse BAC"
1. 1021
/misc_feature
/notes="assembly_fragment
vector end:SP6
vector side:left"
1122_1606
/misc_feature
/notes="assembly_fragment"
1707_2355
/misc_feature
/notes="assembly_fragment"
2456_3109
/misc_feature
/notes="assembly_fragment"
3210_4151
/misc_feature
/notes="assembly_fragment"
4252_4930
/misc_feature
/notes="assembly_fragment"
5031_6123
/misc_feature
/notes="assembly_fragment"
6224_6989
/misc_feature
/notes="assembly_fragment"
7090_7796
/misc_feature
/notes="assembly_fragment"
7897_9525
/misc_feature
/notes="assembly_fragment"
9626_10896
/misc_feature
/notes="assembly_fragment"
10997_13026
/misc_feature
/notes="assembly_fragment"
13127_45197
/misc_feature
/notes="assembly_fragment"

```

```

Query Match 70.9%; Score 23.4; DB 2; Length 194974;
Best Local Similarity 81.8%; Pred. No. 1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTTGTTACTTCTCTCTATTCAGCA 33
Db 49919 GCTTTGTTATTTTCATCTTCATTCAGCA 49951

```

```

RESULT 8
AC091229/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-1A3, *** SEQUENCING IN PROGRESS ***
AC091229
AC091229
AC091229.11 GI:21953795
VERSION
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 336062)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alibrooke,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbada,J., Benton,J., Bimaga,K., Blamendurg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

```

```

COMMENT
AUTHORS
TITLE
JOURNAL

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

AUTHORS
TITLE
JOURNAL

```

```

COMMENT

```

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gerbill,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C.,
 Kravic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louieged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Meigs,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwum,G.,
 Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Pichus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Slason,I.,
 Sodergren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansley,J., Taylor,C., Taylor,T., Teitford,B., Thomas,N., Thomas,S.,
 Uemami,K., Vaequez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.

Unpublished
 2 (bases 1 to 336062)
 Morley,K.C.
 Direct Submission
 Submitted (05-APR-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 336062)
 Morley,K.C.
 Direct Submission
 Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 24, 2002 this sequence version replaced gi:20976487.

Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: TUNA
 Center clone name: CH230-1A3
 Sequencing Statistics
 Sequencing vector: Plasmid;
 Sequencing vector: M13;
 Chemistry: Dye-primer Body: 37% of reads
 Chemistry: Dye-terminator Big Dye: 63% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 151384 bases at least Q40
 Consensus quality: 169228 bases at least Q30
 Consensus quality: 178295 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 185 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1104: contig of 1104 bp in length
 * 1105 1204: gap of unknown length
 * 1205 2243: contig of 1039 bp in length
 * 2244 2343: gap of unknown length
 * 2344 3538: contig of 1195 bp in length
 * 3539 3638: gap of unknown length
 * 3639 4762: contig of 1124 bp in length
 * 4763 4862: gap of unknown length
 * 4863 5906: contig of 1044 bp in length
 * 5907 6006: gap of unknown length
 * 6007 7135: contig of 1129 bp in length
 * 7136 7235: gap of unknown length
 * 7236 8369: contig of 1134 bp in length
 * 8370 8469: gap of unknown length
 * 8470 9516: contig of 1047 bp in length
 * 9517 9616: gap of unknown length
 * 9617 10653: contig of 1037 bp in length
 * 10654 10753: gap of unknown length
 * 10754 11866: contig of 1113 bp in length
 * 11867 11966: gap of unknown length
 * 11967 13111: contig of 1145 bp in length
 * 13112 13211: gap of unknown length
 * 13212 14246: contig of 1035 bp in length
 * 14247 14346: gap of unknown length
 * 14347 15546: contig of 1200 bp in length
 * 15547 15646: gap of unknown length
 * 15647 16761: contig of 1115 bp in length
 * 16762 16861: gap of unknown length
 * 16862 17975: contig of 1114 bp in length
 * 17976 18075: gap of unknown length
 * 18076 19170: contig of 1095 bp in length
 * 19171 19270: gap of unknown length
 * 19271 20440: contig of 1170 bp in length
 * 20441 20540: gap of unknown length
 * 20541 21647: contig of 1107 bp in length
 * 21648 21747: gap of unknown length
 * 21748 22807: contig of 1060 bp in length
 * 22808 22907: gap of unknown length
 * 22908 24027: contig of 1120 bp in length
 * 24028 24127: gap of unknown length
 * 24128 25305: contig of 1178 bp in length
 * 25306 25405: gap of unknown length
 * 25406 26465: contig of 1060 bp in length
 * 26466 26565: gap of unknown length
 * 26566 27794: contig of 1229 bp in length
 * 27795 27894: gap of unknown length
 * 27895 28908: contig of 1014 bp in length
 * 28909 29008: gap of unknown length
 * 29009 30142: contig of 1134 bp in length
 * 30143 30242: gap of unknown length
 * 30243 31433: contig of 1191 bp in length
 * 31434 31533: gap of unknown length
 * 31534 32577: contig of 1044 bp in length
 * 32578 32677: gap of unknown length
 * 32679 33724: contig of 1047 bp in length
 * 33725 33824: gap of unknown length
 * 33825 34898: contig of 1074 bp in length
 * 34899 34998: gap of unknown length
 * 34999 36173: contig of 1175 bp in length
 * 36174 36273: gap of unknown length
 * 36274 37301: contig of 1028 bp in length
 * 37302 37401: gap of unknown length
 * 37402 38832: contig of 1431 bp in length
 * 38833 38932: gap of unknown length
 * 38933 40239: contig of 1307 bp in length
 * 40240 40339: gap of unknown length
 * 40340 41512: contig of 1173 bp in length

41513 41612: gap of unknown length
 * 41613 42959: contig of 1347 bp in length
 * 42960 43059: gap of unknown length
 * 43060 44103: contig of 1044 bp in length
 * 44104 44203: gap of unknown length
 * 44204 45329: contig of 1126 bp in length
 * 45330 45429: gap of unknown length
 * 45430 46682: contig of 1253 bp in length
 * 46683 46782: gap of unknown length
 * 46783 48017: contig of 1235 bp in length
 * 48018 48117: gap of unknown length
 * 48118 49331: contig of 1214 bp in length
 * 49332 49431: gap of unknown length
 * 49432 50543: contig of 1112 bp in length
 * 50544 50643: gap of unknown length
 * 50644 52145: contig of 1502 bp in length
 * 52146 52246: gap of unknown length
 * 52246 53574: contig of 1329 bp in length
 * 53575 53674: gap of unknown length
 * 53675 55265: contig of 1591 bp in length
 * 55266 55365: gap of unknown length
 * 55366 56665: contig of 1300 bp in length
 * 56666 56765: gap of unknown length
 * 56766 57808: contig of 1043 bp in length
 * 57809 57908: gap of unknown length
 * 57909 58981: contig of 1073 bp in length
 * 58982 59081: gap of unknown length
 * 59082 60198: contig of 1117 bp in length
 * 60199 60299: gap of unknown length
 * 60299 61324: contig of 1025 bp in length
 * 61324 61423: gap of unknown length
 * 61424 63199: contig of 1776 bp in length
 * 63200 63299: gap of unknown length
 * 63300 64397: contig of 1099 bp in length
 * 64398 64497: gap of unknown length
 * 64498 65550: contig of 1053 bp in length
 * 65551 65650: gap of unknown length
 * 65651 67161: contig of 1511 bp in length

Query Match 70.9%; Score 23.4; DB 2; Length 336062;
 Best Local Similarity 81.8%; Pred. No. 96;
 Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 GCCTGTCTACTTTCTCTCTATTCAGCA 33
 Db 245630 GCTCTTCTGCTCTCTCTCTTCTTACCA 245598

RESULT 9

AC091347/c

LOCUS

DEFINITION

AC091347

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 345243)
 Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaralunga, H.C., Are, O.R., Ayale, M., Banks, T.,
 Barbarella, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhray, C., Burck, P., Burkett, C., Burnett, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dahome, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.T., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,


```

* 46542 47953: contig of 1412 bp in length
* 47954 48053: gap of unknown length
* 48054 49285: contig of 1232 bp in length
* 49285 49385: gap of unknown length
* 49385 50659: contig of 1274 bp in length
* 50659 50759: gap of unknown length
* 50759 52545: contig of 1786 bp in length
* 52545 52645: gap of unknown length
* 52645 53911: contig of 1266 bp in length
* 53911 54011: gap of unknown length
* 54011 55564: contig of 1553 bp in length
* 55564 55664: gap of unknown length
* 55664 56920: contig of 1256 bp in length
* 56920 57020: gap of unknown length
* 57020 58310: contig of 1290 bp in length
* 58310 58410: gap of unknown length
* 58410 59528: contig of 1119 bp in length
* 59528 59629: gap of unknown length
* 59629 59630: gap of unknown length
* 59630 61186: contig of 1557 bp in length
* 61186 61287: gap of unknown length
* 61287 62625: contig of 1339 bp in length
* 62625 62725: gap of unknown length
* 62725 64167: contig of 1442 bp in length
* 64167 64267: gap of unknown length
* 64267 66271: contig of 2004 bp in length
* 66271 66371: gap of unknown length
* 66371 68090: contig of 1719 bp in length
* 68090 68190: gap of unknown length
* 68190 70130: contig of 1940 bp in length
* 70130 70230: gap of unknown length
* 70230 71628: contig of 1398 bp in length

```

Query Match Best Local Similarity 70.3%; Score 23.4; DB 2; Length 345243;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTTTACTTCTTCTTCTATTCAGCACA 33
Db 133398 GCTGTGTCGTCTCTCTTCTTATACAGCACA 133366

RESULT 10
AC123042/c 177899 bp DNA linear ROD 13-AUG-2002
LOCUS Mus musculus chromosome 10 clone RP24-545017, complete sequence.
DEFINITION AC123042
VERSION AC123042.2 GI:22213542
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 177899)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 177899)
McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 177899)
McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 13, 2002 this sequence version replaced gi:21217585.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu

----- Project Information -----
Center Project name: M BB0545017

Location/Qualifiers
1. 177899 "Mus musculus"
/db_xref="taxon:10090"
/chromosome="10"
/clone="RP24-545017"

BASE COUNT 52391 a 32822 c 33924 g 58762 t
ORIGIN

Query Match Best Local Similarity 70.3%; Score 23.2; DB 10; Length 177899;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGTGTTTACTTCTTCTTCTATTCACA 29
Db 144939 CTGTGTTTCTTCTTCTTCTTCACTCA 144912

RESULT 11
AL391706/c 60921 bp DNA linear PRI 15-NOV-2001
LOCUS Human DNA sequence from clone RP11-431K17 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL391706
VERSION AL391706.9 GI:16972964
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 60921)
AUTHORS Hammond, S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:15793440.

COMMENT During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
RP11-431K17 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-431K17. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-445P19 is at 58922 in this
sequence. The true right end of clone RP11-108M21 is at 2000 in
this sequence.

FEATURES Location/Qualifiers

source 1. .60921
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-431K17"
/clone_1b="RP11-11.2"
BASE COUNT 19409 a 11328 c 10604 g 19580 t
ORIGIN

Query Match 69.7%; Score 23; DB 9; Length 60921;
Best Local Similarity 83.9%; Pred. No. 1.7e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 38675 TGCTGACTTTCTCTCTCTATTCCTCCCA 38645

RESULT 12
AC124546 224860 bp DNA linear HTG 11-AUG-2002
LOCUS Mus musculus chromosome UNK clone RP23-273G17, WORKING DRAFT
DEFINITION
SEQUENCE, 2 unordered pieces.
AC124546
VERSION AC124546.2 GI:22203919
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 224860)
TITLE McPherson,J.D. and Waterston,R.H.
JOURNAL The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 224860)
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 224860)
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Aug 11, 2002 this sequence version replaced gi:21426675.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0273G17
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 224098 bases at least Q40
Consensus quality: 224340 bases at least Q20
Insert size: 19700; agarose-fp
Insert size: 224760; sum-of-contigs
Quality coverage: 11.52 in Q20 bases; agarose-fp
Quality coverage: 11.90 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7505: contig of 7505 bp in length
* 7506 7605: gap of unknown length
* 7606 224860: contig of 217255 bp in length.
Location/Qualifiers
1. .224860
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-273G17"
1. .7505
/note="assembly_name:Contigs5"
/note="assembly_name:Contigs6"
misc_feature 7506 .224860
misc_feature /note="assembly_name:Contigs6"
BASE COUNT 72803 a 40558 c 41158 g 70219 t 122 others
ORIGIN

Query Match 69.7%; Score 23; DB 2; Length 224860;
Best Local Similarity 83.9%; Pred. No. 1.4e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 21703 CTGTTTGCTTTCTCTCTCTATTCCTACC 21733

RESULT 13
AC124707/c 291588 bp DNA linear HTG 22-JUN-2002
LOCUS Mus musculus chromosome UNK clone RP24-73H16, WORKING DRAFT
DEFINITION
SEQUENCE, 7 unordered pieces.
AC124707/c
VERSION AC124707.1 GI:21427964
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 291588)
TITLE McPherson,J.D. and Waterston,R.H.
JOURNAL The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 291588)
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 291588)
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0073H16
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 286901 bases at least Q40
Consensus quality: 288260 bases at least Q30
Consensus quality: 288883 bases at least Q20
Insert size: 19700; agarose-fp

Insert size: 290988; sum-of-contrigs
Quality coverage: 15.82 in Q20 bases; agarose-fp
Quality coverage: 9.51 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15773: contig of 15773 bp in length
* 15774 15873: gap of unknown length
* 15874 41215: contig of 25342 bp in length
* 41216 59868: gap of unknown length
* 59868 59969: gap of unknown length
* 59969 93088: contig of 33119 bp in length
* 93088 93187: gap of unknown length
* 93187 128533: contig of 35346 bp in length
* 128533 128634: gap of unknown length
* 128634 194926: contig of 66293 bp in length
* 194927 195026: gap of unknown length
* 195027 291588: contig of 96562 bp in length.

FEATURES

source

1. 291588
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-73H16"

misc_feature
/note="assembly_name:Contig12"
15874..41215
/note="assembly_name:Contig13"
41316..59868
/note="assembly_name:Contig14"
59969..93087
/note="assembly_name:Contig15"
93188..128533
/note="assembly_name:Contig16"
128634..194926
/note="assembly_name:Contig17"
195027..291588
/note="assembly_name:Contig18"
misc_feature
misc_feature
misc_feature
BASE COUNT 92191 a 54654 c 54286 g 89797 t 660 others
ORIGIN

Query Match 69.7%; Score 23; DB 2; Length 291588;
Best Local Similarity 83.9%; Pred. No. 1.4e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTTCTTATTCAGCC 32

Db 141855 CTGTTTGTCTTCTCTCTATTCACCC 141825

RESULT 14

AL731717 234900 bp DNA linear ROD 24-JUL-2002
LOCUS Mouse DNA sequence from clone RP23-92C4 on chromosome X, complete
DEFINITION
ACCESSION AL731717
VERSION AL731717.10 GI:21955600
KEYWORDS
SOURCE
ORGANISM

Mus musculus.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 234900)
AUTHORS Brown,J.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 25, 2002 this sequence version replaced gi:21690011.
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at:
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-92C4 is from the RP23-Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR::PBACE3.6.

FEATURES

source

1. 234900
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-92C4"
/clone_1fb="RP23-92C4"

BASE COUNT 72446 a 45634 c 45215 g 71605 t
ORIGIN

Query Match 68.5%; Score 22.6; DB 10; Length 234900;
Best Local Similarity 86.2%; Pred. No. 2e+02;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTGTTGTTACTTTCTTCTTATTCGA 29

Db 48695 GCTGTTGTTACTTTCTTCTTATTCGA 48723

RESULT 15

HSA340055/c 664 bp DNA linear PRI 18-JUL-2002
LOCUS Homo sapiens genomic sequence surrounding Not1 site, clone
DEFINITION
ACCESSION AJ340055
VERSION AJ340055.1 GI:15884473
KEYWORDS
SOURCE

Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 664)
AUTHORS Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvaeha,S.M., Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muzvenko,O.V., Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I., Kiselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.

NOT1 flanking sequences: a tool for gene discovery and verification of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
JOURNAL MEDLINE
PUBMED 22131767
REFERENCE 2 (bases 1 to 664)

AUTHORS Zabarovsky, B.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorigenesis Centre,
 Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
 Sweden

FEATURES
 SOURCE Location/Qualifiers
 1. .664
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NR5-1122C"

BASE COUNT 231 a 144 c 148 g 141 t
 ORIGIN

Query Match 67.9%; Score 22.4; DB 9; Length 664;
 Best Local Similarity 81.2%; Pred. No. 5.3e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTTCTTATTCAGCCA 33
 |||||
 Db 419 CTGTTGTTAGTTTCTTATTCGTTTCAGTCA 368

Search completed: May 11, 2003, 06:12:35
 Job time : 604.519 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 64.1805 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33
Sequence: 1 gcgtgtctactcttctctctctcaccgcca 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq.101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	33	21	AAA6850
2	22.4	67.9	569	22	AAK72229
3	22.4	67.9	47319	22	AAK4813
4	22.4	67.9	47319	22	AAK72230
5	22.4	67.9	236303	22	AAK11614
6	22	66.7	458	21	AAK33567
7	22	66.7	13079	22	ABA18638
8	22	66.7	25967	22	ABA18639
9	21.6	65.5	144	22	ABA75350

10	21.6	65.5	144	22	ABA39992	Probe #18458 for g
11	21.6	65.5	144	22	AAK23906	Human brain expres
12	21.6	65.5	144	22	AAK49980	Human bone marrow
13	21.6	65.5	144	22	AAI27077	Probe #17010 for g
14	21.6	65.5	568	22	ABA62842	Human foetal liver
15	21.6	65.5	568	22	ABA30124	Human bone marrow
16	21.6	65.5	568	22	AAK1243	Human brain expres
17	21.6	65.5	568	22	AAK37042	Human bone marrow
18	21.6	65.5	568	22	AAI17884	Probe #7817 for ge
19	21.6	65.5	1057	22	ABK35902	CDNA sequence #293
20	21.4	64.8	5322	22	AAI37484	Human musculoskele
21	21.4	64.8	10644	23	ABU03638	Drosophila melanog
22	21.4	64.8	30568	22	AAI37486	Human musculoskele
23	21.2	64.2	242	20	AAV89996	EST clone CW802.
24	21.2	64.2	1904	22	AAK05148	Human brain expres
25	21.2	64.2	2720	24	ABK10802	Human nucleosomal
26	21.2	64.2	3699	24	ABK10805	DNA encoding human
27	20.8	63.0	222	22	AAK17163	Human brain expres
28	20.8	63.0	1251	24	ABK72973	Bacillus lichenifo
29	20.8	63.0	2312	22	AAK70886	Human immune/haema
30	20.8	63.0	2745	24	ABK84523	Human CDNA differe
31	20.8	63.0	2782	21	AAK52632	Eosinophil activat
32	20.6	62.4	2823	23	ABU25094	Drosophila melanog
33	20.4	61.8	170	15	AAQ76837	Human genome fragm
34	20.4	61.8	1224	23	AAK87626	DNA encoding novel
35	20.4	61.8	1251	22	ABA07827	Human ovarian and
36	20.4	61.8	1251	22	AAU03631	Human reproductive
37	20.4	61.8	112190	22	AAH44801	Human GPCR protein
38	20.2	61.2	693	21	AAU16256	Human prostate can
39	20.2	61.2	1140	21	AAK90085	Human negative elo
40	20.2	61.2	1286	23	AAK85199	DNA encoding novel
41	20.2	61.2	1301	24	ABN95636	Gene #2134 used to
42	20.2	61.2	3758	17	AAU44318	Human cytokine HCC
43	20.2	61.2	4037	24	ABN97447	Gene #3945 used to
44	20.2	61.2	4037	24	ABU63716	Breast cancer rela
45	20.2	61.2	4959	22	AAH81756	Human differential

ALIGNMENTS

RESULT 1	
AAA96850	AAA96850 standard; DNA; 33 BP.
XX	XX
AC	AAA96850;
XX	XX
DT	19-FEB-2001 (first entry)
XX	XX
DE	Guide desoxynucleotide building block G2.
XX	XX
KM	Promoter; intergenic region; Comelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KM	transgenic plant; ss.
XX	XX
OS	Synthetic.
XX	XX
PN	WO200058485-A1.
XX	XX
PD	05-OCT-2000.
XX	XX
PF	29-MAR-2000; 2000MO-IB00370.
XX	XX
PR	29-MAR-1999; 99FR-0003925.
XX	XX
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	XX
PI	Rance I, Gruber V, Theisen M;
DR	WPI; 2000-647238/62.
XX	XX
PT	Chimeric expression promoter for transgenic plant production, comprises

PT Sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 24; 91pp; English.
XX
CC The present sequence represents a guide deoxynucleotide building
CC block, which was used to link directional deoxynucleotide building
CC blocks during construction of chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comeline yellow mottle virus, and the second plant promoter originates
CC from the Casava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
CC
XX
SO Sequence 33 BP; 4 A; 9 C; 4 G; 16 T; 0 other;
Query Match 100.0%; Score 33; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GCTGTTACTTCTCTCTATTCAGCA 33
Db 1 GCTGTTACTTCTCTCTATTCAGCA 33
RESULT 2
AAK72229/c
ID AAK72229 standard; DNA; 569 BP.
XX
AC AAK72229;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27041.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
PD
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0230963.
PR 26-JUL-2000; 2000US-0230964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225799.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234474.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236379.
PR 02-OCT-2000; 2000US-0236602.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240360.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

XX	08-NOV-2000	2000US-0246525	PR
XX	08-NOV-2000	2000US-0246524	PR
XX	08-NOV-2000	2000US-0246525	PR
XX	08-NOV-2000	2000US-0246526	PR
XX	08-NOV-2000	2000US-0246527	PR
XX	08-NOV-2000	2000US-0246528	PR
XX	08-NOV-2000	2000US-0246532	PR
XX	08-NOV-2000	2000US-0246539	PR
XX	08-NOV-2000	2000US-0246602	PR
XX	08-NOV-2000	2000US-0246610	PR
XX	08-NOV-2000	2000US-0246611	PR
XX	08-NOV-2000	2000US-0246613	PR
XX	17-NOV-2000	2000US-0249207	PR
XX	17-NOV-2000	2000US-0249208	PR
XX	17-NOV-2000	2000US-0249209	PR
XX	17-NOV-2000	2000US-0249210	PR
XX	17-NOV-2000	2000US-0249211	PR
XX	17-NOV-2000	2000US-0249212	PR
XX	17-NOV-2000	2000US-0249213	PR
XX	17-NOV-2000	2000US-0249214	PR
XX	17-NOV-2000	2000US-0249215	PR
XX	17-NOV-2000	2000US-0249216	PR
XX	17-NOV-2000	2000US-0249217	PR
XX	17-NOV-2000	2000US-0249218	PR
XX	17-NOV-2000	2000US-0249244	PR
XX	17-NOV-2000	2000US-0249245	PR
XX	17-NOV-2000	2000US-0249264	PR
XX	17-NOV-2000	2000US-0249265	PR
XX	17-NOV-2000	2000US-0249297	PR
XX	17-NOV-2000	2000US-0249299	PR
XX	17-NOV-2000	2000US-0249300	PR
XX	01-DEC-2000	2000US-0250160	PR
XX	01-DEC-2000	2000US-0250391	PR
XX	05-DEC-2000	2000US-0251030	PR
XX	05-DEC-2000	2000US-0251088	PR
XX	05-DEC-2000	2000US-0251719	PR
XX	06-DEC-2000	2000US-0251479	PR
XX	08-DEC-2000	2000US-0251856	PR
XX	08-DEC-2000	2000US-0251868	PR
XX	08-DEC-2000	2000US-0251869	PR
XX	08-DEC-2000	2000US-0251989	PR
XX	11-DEC-2000	2000US-0254097	PR
XX	05-JAN-2001	2001US-0259678	PR
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI, 2001-483426/52.		
XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
XX	useful for preventing, diagnosing and/or treating cancers and		
XX	metastasis -		
XX	Disclosure; SEQ ID NO 27041; 3071pp + Sequence Listing; English.		
XX	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)		
XX	antigen acid sequences given in AAM82170 to AAM91921. (I) have cytostratic		
XX	activity, and can be used in gene therapy and vaccine production. (I)		
XX	proteins and polynucleotides may be used in the prevention, diagnosis and		
XX	treatment of diseases associated with inappropriate (I) expression. For		
XX	example, they may be used to treat disorders associated with decreased		
XX	expression by rectifying mutations or deletions in a patient's genome		
XX	that affect the activity of (I) by expressing inactive proteins or to		
XX	supplement the patient's own production of (I). Additionally, (I)		
XX	polynucleotides may be used to produce the secreted (I), by inserting		
XX	the nucleic acids may be used to produce the secreted (I), by inserting		
XX	the protein. (I) proteins and polynucleotides may be used to prevent,		
XX	diagnose and treat immune/hematopoietic-related diseases, especially		
XX	cancers and cancer metastases of hematopoietic-derived cells. AAK4703		
XX	to AAK87694 represent human immune/hematopoietic antigen genomic		
XX	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		
XX	represent sequences used in the exemplification of the present invention.		

Seq	Sequence	569 BP; 162 A; 83 C; 156 G; 168 T; 0 other;
Query Match		67.9%; Score 22.4; DB 22; Length 569;
Best Local Similarity	81.2%;	Pred. No. 91;
Matches	26; Conservative	0; Mismatches 6; Indels 0; Gaps 0;
Oy	1	GCTGTGTTACTTTCCTCTCATTCAGCC 32
Db	551	GCTGATGCTACTTACCTTCTTGTAATTCAGCC 520

RESULT 3	
ID	AAK64813/c
	AAK64813 standard; DNA; 47319 BP.
AC	AAK64813;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19625.
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
FN	WO200157182-A2.
PD	
XX	
PF	09-AUG-2001.
XX	
XX	17-JAN-2001; 2001WO-US01354.
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS Homo sapiens.
PN WO200157182-A2.
XX 09-AUG-2001.
PD 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180668.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218280.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27042; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 47319 BP; 13708 A; 8453 C; 9720 G; 15438 T; 0 other;
Query Match 67.9%; Score 22.4; DB 22; Length 47319;
Best Local Similarity 81.2%; Pred. NO. 1.2e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 1 GCTGTGTTACTTTCTCTCTATTCAGCC 32
DB 12933 GCTGATGTCCTTACTTCTGATTCACCC 12902
RESULT 5
AAS11614
ID AAS11614 standard; DNA; 236303 BP.
XX
AC AAS11614;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human genomic DNA containing exons 2-17 of the CRIM1 gene.
XX
KM CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
KM neuroprotective; renal; osteopathic; dental; vulnaray; immunogen;
KM antibody; gene therapy; neurodegenerative disease; eye disorder;
KM cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
KM tooth abnormality; wound; ds.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT 33104..33277
FT /*tag= a
FT /number= 2
FT 33278..77746
FT /*tag= b
FT /number= 2
FT 77747..77989
FT /*tag= c
FT /number= 3
FT 77990..79103
FT intron

FT /*tag= d
FT /number= 3
FT 79104..79224
FT /*tag= e
FT /number= 4
FT 79225..101022
FT /*tag= f
FT /number= 4
FT 101023..101144
FT /*tag= g
FT /number= 5
FT 101145..113377
FT /*tag= h
FT /number= 5
FT 113378..113560
FT /*tag= i
FT /number= 6
FT 113561..115985
FT /*tag= j
FT /number= 6
FT 115986..116183
FT /*tag= k
FT /number= 7
FT 116184..135707
FT /*tag= l
FT /number= 7
FT 135708..135836
FT /*tag= m
FT /number= 8
FT 135837..146471
FT /*tag= m
FT /number= 8
FT 146472..146628
FT /*tag= o
FT /number= 9
FT 146629..148761
FT /*tag= p
FT /number= 9
FT 148762..148883
FT /*tag= q
FT /number= 10
FT 148884..150044
FT /*tag= r
FT /number= 10
FT 150045..150254
FT /*tag= s
FT /number= 11
FT 150255..153815
FT /*tag= t
FT /number= 11
FT 153816..154031
FT /*tag= u
FT /number= 12
FT 154032..158580
FT /*tag= v
FT /number= 12
FT 158581..158802
FT /*tag= w
FT /number= 13
FT 158803..173982
FT /*tag= x
FT /number= 13
FT 173983..174177
FT /*tag= y
FT /number= 14
FT 174178..181006
FT /*tag= z
FT /number= 14
FT 181007..181129
FT /*tag= aa
FT /number= 15
FT 181130..183612
FT /*tag= ab
FT intron

FT exon /number= 15
183613..183800
/*cag= ac
FT intron /number= 16
183801..185152
/*cag= ad
FT exon /number= 16
185153..187765
/*cag= ae
FT /number= 17
XX MO200138519-A1.
XX 31-MAY-2001.
XX 24-NOV-2000; 2000MO-AU01435.
XX 26-NOV-1999; 99AU-0004348.
XX (UYOU) UNIV QUEENSLAND.
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX WPI; 2001-343951/36.
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX useful for preventing, diagnosing and treating e.g. eye disease,
XX especially cataract formation -
XX Claim 4; Fig 3; 169pp; English.
XX The invention relates to nucleic acids from human chromosome 2p21-16.3
XX and the encoded peptide (and mouse and chicken orthologues) that
XX comprises a RGCGRP group, an insulin-like growth factor binding protein
XX (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
XX and a transmembrane domain. The protein, e.g. CRIM1, interacts with
XX peptides of the transforming growth factor superfamily. A composition
XX comprising an expression construct comprising the nucleic acids of the
XX invention or a mimetic which antagonises or mimics an activity of a CRIM1
XX polypeptide may be used in a method for modulating the biological
XX activity of a polypeptide of the bone morphogenic protein (BMP) family.
XX In this way they may be used to prevent or treat an eye disease,
XX especially cataract formation. They may also be used to treat
XX neurodegenerative diseases, renal and kidney disease, bone and tooth
XX abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
XX gene therapy by using antibodies directed against CRIM1 polypeptides.
XX The present sequence is a Human genomic DNA containing exons 2-17 of
XX the CRIM1 gene.
XX Sequence 236303 BP; 70199 A; 46544 C; 47996 G; 71563 T; 1 other;
XX
XX Query Match 67.9%; Score 22.4; DB 22; Length 236303;
XX Best Local Similarity 81.2%; Pred. No. 1.4e+02;
XX Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 1 GCTGTGTACTTCTTCTTCAATCCAGCC 32
XX |||||
Db 214710 GCTGATGCTACTTACTTCTTATTCACAC 214741
XX
XX RESULT 6
XX AAC43567
XX ID AAC43567 standard; DNA; 458 BP.
XX AAC43567;
XX
XX 17-OCT-2000 (first entry)
XX
XX Zea mays DNA fragment SEQ ID NO: 39699.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic;
XX pathway; promoter; termination sequence; corn; ss.

XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139452.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139464.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-013624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147058.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 22; DB 21; Length 458;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCTGTGTTACTTCTTCTTCTTATTCAG 30
Db 114 GCTTTGTACTACTCTTCTTCTTCTTCAG 143

RESULT 7
ID ABA18638/c
ABAB18638 standard; DNA; 13079 BP.

AC ABA18638;

XX 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 10969.

XX Human; noctropic; neuroprotective; cytoskeletal; dermatological; viroicide;
KW Immunosuppressive; antihistaminic; anti-HIV; antibacterial; vulnecary;
KW antiparkinsonian; antistickling; antianaemic; antitachytic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antineutrophilic;
KW antileukemic; antidiabetic; antileucic; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; de.
XX Homo sapiens.
OS

PN WC200159063-A2.
 XX 16-AUG-2001.
 PD
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217949.
 PR 26-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229345.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 06-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239395.
 PR 13-OCT-2000; 2000US-0239397.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 11-DEC-2000; 2000US-0254997.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen Ca, Barash SC, Ruben SM;
 PI

XX DR WPI; 2001-541565/60.
 XX XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides.
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure; SEQ ID NO 10969; 1701bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
 CC (ABAI1678-ABAI18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 13079 BP; 4243 A; 2205 C; 2671 G; 3960 T; 0 other;
 Query Match 66.7%; Score 22; DB 22; Length 13079;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 GCTGTGTTACTTTCTCTCTATTCAG 30
 Db 11911 GGTGTGTTACTTTCTCTCTATTCAG 11882
 RESULT 8
 ABAI18639/C
 ID ABAI18639 standard; DNA; 25967 BP.
 XX AC ABAI18639;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 10970.
 XX
 KW Human; goitrotic; neuroprotective; cytostatic; dermatological; virocidic;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;
 KW antiparkinsonian; antichilling; antianemic; antiallergic; cancer;
 KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200159063-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216680.
 PR 07-JUL-2000; 2000US-0216887.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 25-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-541565/60.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 XX
 PS Disclosure; SEQ ID NO 10970; 1701bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
 CC (ABAI1678-ABAI1800) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 25967 BP; 8245 A; 4194 C; 4708 G; 8820 T; 0 other;
 Query Match 66.7%; Score 22; DB 22; Length 25967;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 11919 GGTGTGTACATTCTTTCTATTCAG 11890
 GY 1 GCTGTGTACTTTCTTTCTATTCAG 30
 ID ABA75350 standard; DNA; 144 BP.
 AC ABA75350;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #23655.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001MO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -
 XX
 XX
 PS Claim 4; SEQ ID NO 23655; 639bp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
SQ Query Match 65.5%; Score 21.6; DB 22; Length 144;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 29 TTTGTCTTCTTCTTCTTCTCTCTAG 56

RESULT 10
ABA39992
ID ABA39992 standard; DNA; 144 BP.
XX ABA39992;
AC 23-JAN-2002 (first entry)
XX
DE Probe #18458 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KM cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX Homo sapiens.
OS
XX WO200157274-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488899/53.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 18458; 530bp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
SQ Query Match 65.5%; Score 21.6; DB 22; Length 144;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGTGTACTTTCTTCTTCTATTCAG 30

DB 29 TTTGTCTTCTTCTTCTTCTCTAG 56
|||||

RESULT 11
AAK3906
ID AAK3906 standard; DNA; 144 BP.
XX AAK3906;
AC 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 23897.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 23897; 650bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
SQ Query Match 65.5%; Score 21.6; DB 22; Length 144;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 29 TTTGTCTTCTTCTTCTTCTCTAG 56

RESULT 12
AAK4980
ID AAK4980 standard; DNA; 144 BP.
XX
XX AAK4980;
AC 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 24537.

```

XX Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
OS Homo sapiens.
PN MO200157276-A2.
XX
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PP
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488990/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 24537; 658pp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
SQ
XX
XX Query Match 65.5%; Score 21.6; DB 22; Length 144;
XX Best Local Similarity 85.7%; Pred. No. 1.6e+02;
XX Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY
XX 3 TGTGTACTTTCTTCCTTCATTCAG 30
XX |||||
XX 29 TTTTGTCTTTCTTCTCTCTCTAG 56
XX
XX RESULT 13
XX AA127077
XX ID AA127077 standard; DNA; 144 BP.
XX
XX AA127077;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #17010 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX MO200157278-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX

```

```

PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
XX      04-OCT-2000; 2000GB-0024263.
PA      (MOLE-) MOLECULAR DYNAMICS INC.
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
DR      WPI; 2001-488901/53.
XX      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human cervical epithelial cells -
XX      Claim 25; SEQ ID No 17010; 487bp; English.
PS      The present invention relates to human single exon nucleic acid probes
CC      (SENPs). The present sequence is one such probe. The SENPs are derived
CC      from human HeLa cells. The SENPs can be used to produce a single exon
CC      microarray, which can be used for measuring human gene expression in a
CC      sample derived from human cervical epithelial cells. By measuring gene
CC      expression, the probes are therefore useful in grading and/or staging
CC      of diseases of the cervix, notably cervical cancer.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
SQ
Query Match          65.5%; Score 21.6; DB 22; Length 144;
Best Local Similarity 85.7%; Pred. NO. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      3 TGTGTTACTTTCCTCTTCCTCATTCAG 30
Db      29 TTGGTTCCTTCTCCTCTCTCTCTTAG 56
RESULT 14
ABA62842
ID      ABA62842 standard; DNA; 568 BP.
XX      ABA62842;
AC      ABA62842;
DT      01-FEB-2002 (first entry)
XX      Human foetal liver single exon nucleic acid probe #11147.
DE      Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX      Homo sapiens.
OS      Homo sapiens.
XX      WO200157277-A2.
PN      09-AUG-2001.
XX      30-JAN-2001; 2001WO-US00669.
PF      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
XX      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
XX      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
XX      04-OCT-2000; 2000GB-0024263.
PA      (MOLE-) MOLECULAR DYNAMICS INC.
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
DR      WPI; 2001-483447/52.
XX      Human genome-derived single exon nucleic acid probes useful for
LT
```

PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 11147; 639bp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 568 BP; 196 A; 95 C; 69 G; 208 T; 0 other;

Query Match 65.5%; Score 21.6; DB 22; Length 568;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGTGTTACTTTCTCTCTATTCAG 30
Db 97 TTTGTTCTTTCTCTCTCTCTAG 124

RESULT 15
ABA30124
ID ABA30124 standard; DNA; 568 BP.
XX
AC ABA30124;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #8590 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PR 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 8590; 530bp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 568 BP; 196 A; 95 C; 69 G; 208 T; 0 other;

Query Match 65.5%; Score 21.6; DB 22; Length 568;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGTGTTACTTTCTCTCTATTCAG 30
Db 97 TTTGTTCTTTCTCTCTCTAG 124

Search completed: May 11, 2003, 03:06:50
Job time : 108.18 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 / Search time 13.231 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33

Sequence: 1 gctgtgtctactcttctctctcaccagca 33

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Issued_Patents_NA.*
- 2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.4	61.8	26173	4 US-09-453-702B-69	Sequence 69, App1
C 2	19.8	60.0	4145	4 US-09-302-620B-82	Sequence 82, App1
C 3	19.8	60.0	4206	4 US-09-302-620B-81	Sequence 81, App1
C 4	19.6	59.4	2623	2 US-08-973-675-1	Sequence 59, App1
C 5	19.6	59.4	14855	2 US-08-687-080-59	Sequence 15, App1
C 6	19	57.6	1314	1 US-07-662-005A-15	Sequence 2, App1
C 7	19	57.6	1351	4 US-09-002-114-2	Sequence 7, App1
C 8	18.8	57.0	594	4 US-09-687-477-7	Sequence 11, App1
C 9	18.8	57.0	594	4 US-09-687-477-11	Sequence 13, App1
C 10	18.8	57.0	594	4 US-09-687-477-13	Sequence 15, App1
C 11	18.8	57.0	594	4 US-09-687-477-15	Sequence 6, App1
C 12	18.8	57.0	45546	4 US-09-146-053-6	Sequence 1709, App1
C 13	18.6	56.4	1191	4 US-09-134-001C-1709	Sequence 3, App1
C 14	18.6	56.4	2196	4 US-08-472-217-3	Sequence 14, App1
C 15	18.6	56.4	2196	3 US-08-760-534A-3	Sequence 16, App1
C 16	18.6	56.4	7350	1 US-07-865-662F-14	Sequence 1, App1
C 17	18.6	56.4	7350	4 US-08-374-219B-16	Sequence 12, App1
C 18	18.6	56.4	168575	4 US-09-426-280-1	Sequence 5, App1
C 19	18.4	55.8	1284	4 US-09-134-001C-212	Sequence 15, App1
C 20	18.4	55.8	2543	4 US-09-307-143-5	Sequence 14, App1
C 21	18.4	55.8	3387	1 US-08-468-557-1	Sequence 14, App1
C 22	18.4	55.8	17327	1 US-07-906-871-15	Sequence 14, App1
C 23	18.2	55.2	223	4 US-08-205-697A-14	Sequence 14, App1
C 24	18.2	55.2	223	4 US-08-702-525-14	Sequence 14, App1
C 25	18.2	55.2	223	5 PCT-US95-02576-14	Sequence 14, App1
C 26	18.2	55.2	301	4 US-09-605-785-234	Sequence 234, App1
C 27	18.2	55.2	301	4 US-09-439-313-234	Sequence 234, App1

28	18.2	55.2	301	4 US-09-352-616A-234	Sequence 234, App
29	18.2	55.2	301	4 US-09-232-149A-234	Sequence 234, App
30	18.2	55.2	503	4 US-09-280-116-203	Sequence 203, App
31	18.2	55.2	589	4 US-09-328-111-546	Sequence 546, App
32	18.2	55.2	659	2 US-08-718-538-4	Sequence 4, App1
33	18.2	55.2	890	2 US-08-319-376-1	Sequence 1, App1
34	18.2	55.2	1005	2 US-08-701-935-4	Sequence 4, App1
35	18.2	55.2	1005	3 US-09-134-591-4	Sequence 12, App1
36	18.2	55.2	1261	4 US-08-205-697A-12	Sequence 12, App1
37	18.2	55.2	1261	4 US-08-702-525-12	Sequence 12, App1
38	18.2	55.2	1261	5 PCT-US95-02576-12	Sequence 12, App1
39	18.2	55.2	1509	4 US-09-149-476-179	Sequence 179, App
40	18.2	55.2	2896	1 US-08-441-430-31	Sequence 31, App1
41	18.2	55.2	2995	1 US-08-441-430-32	Sequence 32, App1
42	18.2	55.2	3387	1 US-08-064-271-11	Sequence 29, App1
43	18.2	55.2	3387	2 US-08-627-525C-29	Sequence 29, App1
44	18.2	55.2	3387	3 US-08-530-589A-19	Sequence 19, App1
45	18.2	55.2	3387	4 US-09-599-781-19	Sequence 19, App1

ALIGNMENTS

RESULT 1
US-09-453-702B-69/c
Sequence 69, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Nicole T.
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453, 702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-Dec-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 26173
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-09-453-702B-69
Query Match 61.8%; Score 20.4; DB 4; Length 26173;
Best Local Similarity 80.0%; Pred. No. 58;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 4 GTGTACTTCTCTCTATTCAGCA 33

Db 8207 GTGTACTTCTCTCTATTCAGCA 8178

RESULT 2

US-09-302-6208-82/c

Sequence 82, Application US/093026208

Patent No. 6331420

GENERAL INFORMATION:

APPLICANT: Wilson, C. Ron

APPLICANT: Craft, David L.

APPLICANT: Birch, Dudley

APPLICANT: Eschro, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450

TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA

TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS

FILE REFERENCE: 1010-16, seq

CURRENT APPLICATION NUMBER: US/09/302,6208

NUMBER OF SEQ ID NOS: 109

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 82

LENGTH: 4145

TYPE: DNA

ORGANISM: Candida tropicalis

US-09-302-6208-82

Query Match

Best Local Similarity 77.4%; Score 19.8; DB 4; Length 4145;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TGTGTACTTCTCTCTATTCAGCA 33

Db 1207 TGTGTACTTCTCTCTATTCAGCA 1177

RESULT 3

US-09-302-6208-81/c

Sequence 81, Application US/093026208

Patent No. 6331420

GENERAL INFORMATION:

APPLICANT: Wilson, C. Ron

APPLICANT: Craft, David L.

APPLICANT: Birch, Dudley

APPLICANT: Eschro, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450

TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA

TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS

FILE REFERENCE: 1010-16, seq

CURRENT APPLICATION NUMBER: US/09/302,6208

NUMBER OF SEQ ID NOS: 109

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 81

LENGTH: 4206

TYPE: DNA

ORGANISM: Candida tropicalis

US-09-302-6208-81

ORGANISM: Candida tropicalis

US-09-302-6208-81

Query Match

Best Local Similarity 77.4%; Score 19.8; DB 4; Length 4206;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TGTGTACTTCTCTCTATTCAGCA 33

Db 1180 TGTGTACTTCTCTCTATTCAGCA 1150

RESULT 4

US-08-973-675-1/c

Sequence 1, Application US/08973675

Patent No. 5985283

GENERAL INFORMATION:

APPLICANT: HATBOER, GUNS

APPLICANT: BERNARDS, RENE

TITLE OF INVENTION: ADENOVIRUS E1A-ASSOCIATED PROTEIN BS69,

TITLE OF INVENTION: INHIBITOR OF E1A-TRANSACTIVATION

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,675

FILING DATE: 12-DEC-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: WILSON, MARY J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 620-31

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2623 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 150..1835

US-08-973-675-1

Query Match

Best Local Similarity 84.4%; Score 19.6; DB 2; Length 2623;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 5 TTGTACTTCTCTCTATTCAG 30

Db 1562 TTGTACTTCTCTCTATTCAG 1537

RESULT 5

US-08-687-080-59

Sequence 59, Application US/08687080

Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 14855 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC
INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-59

Query Match 59.4%; Score 19.6; DB 2; Length 14855;
Best Local Similarity 84.6%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY 8 TTACTTTCTTCTTCTATTCAGCCA 33
DB 5760 TTACTTTCTTCTTCTGCTCAGCCA 5785

RESULT 6
US-07-662-005A-15
Sequence 15, Application US/07662005A
Patent No. 5246838
GENERAL INFORMATION:
APPLICANT: Van Dijk, Jan M.
APPLICANT: Smith, Hilde B.
APPLICANT: Bron, Sierd
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: IMPROVED PROCESSING OF PROTEINS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/662,005A
FILING DATE: 19910228
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152002300
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-662-005A-15

Query Match 57.6%; Score 19; DB 1; Length 1314;
Best Local Similarity 81.5%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CY 4 GTTGTTACTTTCTTCTTCTATTCAG 30
DB 554 GTAGTTACTTCTTCTTCTTCTTACAG 580

RESULT 7
US-09-002-114-2/c
Sequence 2, Application US/09002114
Patent No. 6274720
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN PREPRONEUROTENSIN/NEUROMEDIN N
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,114
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0450 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-002-114-2

Query Match 57.6%; Score 19; DB 4; Length 1351;
Best Local Similarity 81.5%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTCTCTATTC 28

Db 1306 CTGTTTACTTCTCTTTCTTTTC 1280

RESULT 8

US-09-687-477-7
; Sequence 7, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Chum Salmon
US-09-687-477-7

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCATTATCTGCTGTTTCCAGC 148

RESULT 9

US-09-687-477-11
; Sequence 11, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 594
; TYPE: DNA
; ORGANISM: King Salmon
US-09-687-477-11

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCATTATCTGCTGTTTCCAGC 148

RESULT 10
US-09-687-477-13
; Sequence 13, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Pink Salmon
US-09-687-477-13

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCATTATCTGCTGTTTCCAGC 148

RESULT 11

US-09-687-477-15
; Sequence 15, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Sockeye Salmon
US-09-687-477-15

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCATTATCTGCTGTTTCCAGC 148

RESULT 12

US-09-146-053-6/c
; Sequence 6, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.

```

? TITLE OF INVENTION: Human Aminopeptidase P Gene
?
? FILE REFERENCE: MCG103
?
? CURRENT APPLICATION NUMBER: US/09/146,053A
?
? CURRENT FILING DATE: 1998-09-02
?
? EARLIER APPLICATION NUMBER: 60/057,854
?
? EARLIER FILING DATE: 1997-09-02
?
? NUMBER OF SEQ ID NOS: 7
?
? SOFTWARE: PatentIn Ver. 2.0
?
? SEQ ID NO 6
?
? LENGTH: 45546
?
? TYPE: DNA
?
? ORGANISM: Homo sapiens
?
US-09-146-053-6

```

Query Match	57.0%	Score 18.8;	DB 4;	Length 45546;
Best Local Similarity	76.7%	Pred. No. 2.4e+02;		
.Matches 23; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;

```

QY      3  TGTGTACTTTCTCTCTATCCAGCC  32
          ||| ||| ||| ||| ||| ||| ||| |||
Db 10468 TGACTTTAGTTTCATCTCTATCCCTGCC 10439

```

```

RESULT 13
US-09-134-001C-1709/c
: Sequence 1709, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1709
: LENGTH: 1191
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1709

```

Query Match	56.4%	Score 18.6	DB 4	Length 1191
Best Local Similarity	84.0%	Pred. No. 1.7e+02		
Matches 21	Conservative 0	Mismatches 4	Indels 0	Gaps 0

QY 2 CTGTTGTTACTTTCTTCTTCTATT 26
||| ||| ||| ||| ||| ||| |||
Db 837 CTGTTCTTTCTTATCTTCGTCTATT 813

RESULT 14
US-08-472-217-3
Sequence 3, Application US/08472217
Patent No. 5726058
GENERAL INFORMATION:
APPLICANT: Aalanen-Kurki, Leena
APPLICANT: Aalunen, Petri
APPLICANT: Jaakkola, Panu
APPLICANT: Jalkanen, Markku
APPLICANT: Lepp, Sirpa
APPLICANT: Maki, Markku
APPLICANT: Vahnen, Tapani
APPLICANT: Wirt, Anni
TITLE OF INVENTION: Syndecan Stimulation Of Cellular
NUMBER OF INVENTION: Differentiation
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,217
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,186
FILING DATE: 07-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,427
FILING DATE: 01-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 1102.0050003
TELECOMMUNICATIONS INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ. ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2196 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-472-217-3

Query Match	56.4%	Score 18.6	DB 1	Length 2196
Best Local Similarity	72.7%	Pred. NO. 1.8e+02		
Matches 24	Conservative 0	Mismatches 9	Indels 0	Gaps 0

OY

1 GCTGTTGTACTTTCCTTCCTCATTCCAGGCCA 33
| | | | | | | | | | | | | |
Db 224 GTTGTTGCCTCTTTCCTCCTCTTGGTTCAGCCA 256
| | | | | | | | | | | | | |

RESULT 15
 US-08-760-534A-3
 Sequence 3, Application US/08760534A
 Patent No. 6017727
 GENERAL INFORMATION:
 APPLICANT: JALKANEN, MARKKU
 APPLICANT: JAAKKOLA, PANU
 APPLICANT: VIHINEN, TAPANI
 TITLE OF INVENTION: SYNDSCAN ENHANCER ELEMENT AND SYNDSCAN
 TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P. L. L. C.
 STREET: 1100 NEW YORK AVENUE, SUITE 600
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: US
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/760,534A
 FILING DATE: 02-DEC-1996
 CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,186
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA: PCT/FI93/00514
; APPLICATION NUMBER: PCT/FI93/00514
; FILING DATE: 01-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CIMBALA, MICHELE A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-760-534A-3

```

```

Query Match      56.4%; Score 18.6; DB 3; Length 2196;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
OY 1 GCTGTTGTTACTTCTCTCTCTATTCAGCA 33
Db 224 GTGTTGTTGTTACTTCTCTCTCTATTCAGCA 256

```

Search completed: May 11, 2003, 03:08:01
 Job time : 27.2331 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 32.6692 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33
Sequence: 1 gctgtgtactcttctctctcaccagca 33

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	33	US-09-963-803-16	Sequence 16, Appl
2	22.4	67.9	36303	US-10-152-724A-24	Sequence 24, Appl
3	21.6	65.5	144	US-09-864-761-25312	Sequence 25312, A
4	21.6	65.5	310	US-09-796-692-6086	Sequence 6086, Ap
5	21.6	65.5	310	US-10-040-862-6086	Sequence 6086, Ap
6	21.6	65.5	408	US-09-796-692-4667	Sequence 4667, Ap
7	21.6	65.5	408	US-09-796-692-5376	Sequence 5376, Ap
8	21.6	65.5	408	US-09-796-692-7106	Sequence 7106, Ap
9	21.6	65.5	408	US-10-040-862-4667	Sequence 4667, Ap
10	21.6	65.5	408	US-10-040-862-5376	Sequence 5376, Ap
11	21.6	65.5	408	US-10-040-862-7106	Sequence 7106, Ap
12	21.6	65.5	499	US-09-783-590-151	Sequence 151, App
13	21.6	65.5	568	US-09-864-761-8590	Sequence 8590, Ap
14	21.6	65.5	1057	US-09-822-846-293	Sequence 8590, Ap
15	21.4	64.8	356	US-09-938-842A-5185	Sequence 5185, Ap
16	21.4	64.8	5322	US-09-764-877-3849	Sequence 3849, Ap
17	21.4	64.8	30568	US-09-764-877-3851	Sequence 3851, Ap
18	21.2	64.2	1904	US-09-864-761-10600	Sequence 10600, A
19	21.2	64.2	2720	US-09-847-665-3	Sequence 3, Appl

20	21.2	64.2	3699	10	US-09-847-665-6	Sequence 6, Appl
21	20.8	63.0	222	10	US-09-864-761-27032	Sequence 27032, A
22	20.8	63.0	1251	10	US-09-974-300-264	Sequence 264, App
23	20.4	61.8	1251	9	US-09-764-891-6319	Sequence 6319, Ap
24	20.4	61.8	26173	9	US-10-114-170-69	Sequence 69, Appl
25	20.2	61.2	442	9	US-09-796-692-8072	Sequence 8072, Ap
26	20.2	61.2	442	9	US-10-040-862-8072	Sequence 8072, Ap
27	20.2	61.2	693	10	US-09-925-300-691	Sequence 691, App
28	20.2	61.2	1301	10	US-09-880-107-2134	Sequence 2134, Ap
29	20.2	61.2	4037	9	US-09-954-531-986	Sequence 986, App
30	20.2	61.2	4037	10	US-09-880-107-3942	Sequence 3942, Ap
31	20.2	61.2	11879	10	US-09-070-927A-181	Sequence 181, App
32	20.2	61.2	32203	9	US-10-091-504-1849	Sequence 1849, Ap
33	20.2	61.2	32203	10	US-09-764-869-1849	Sequence 1849, A
34	20.2	60.6	180	10	US-09-864-761-21254	Sequence 21254, A
35	20.2	60.6	180	10	US-09-864-761-22830	Sequence 22830, A
36	20.2	60.6	456	10	US-09-864-761-4507	Sequence 4507, Ap
37	20.2	60.6	599	10	US-09-864-761-13264	Sequence 13264, A
38	20.2	60.6	1735	10	US-09-070-927A-525	Sequence 525, Appl
39	20.2	60.6	1742	10	US-09-925-300-33	Sequence 33, Appl
40	20.2	60.6	4859	10	US-09-954-456-493	Sequence 493, App
41	20.2	60.6	4859	10	US-09-880-107-1647	Sequence 1647, Ap
42	20.2	60.6	5325	9	US-10-079-854-340	Sequence 340, App
43	20.2	60.6	5325	10	US-09-764-878-340	Sequence 340, App
44	20.2	60.6	5450	12	US-10-044-090-504	Sequence 504, App
45	19.8	60.0	117	10	US-09-969-373-488	Sequence 488, App

ALIGNMENTS

RESULT 1
US-09-963-803-16
Sequence 16, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 33
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Guide desoxynucleotide building
US-09-963-803-16
Query Match 100.0%; Score 33; DB 9; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 GCTGTGTACTCTTCTCTCTCATTCCAGCA 33
OY |||
1 GCTGTGTACTCTTCTCTCTCATTCCAGCA 33
US-10-152-724A-24
Sequence 24, Application US/10152724A
Publication No. US20030082714A1
GENERAL INFORMATION:
APPLICANT: LITTELL, Melissa
APPLICANT: HOLMES, Gregory

APPLICANT: KOLLE, Gabriel
APPLICANT: YAMADA, Toshiya
APPLICANT: GEORGAS, Kylie
APPLICANT: WILKINSON, Lorine
TITLE OF INVENTION: No. US2003008271A1el Nucleic Acid and Polypeptide
FILE REFERENCE: P22378
CURRENT APPLICATION NUMBER: US/10/152,724A
CURRENT FILING DATE: 2002-05-23
PRIOR APPLICATION NUMBER: Australian App No. US2003008271A1PQ 4348
PRIOR FILING DATE: 1999-11-26
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 36303
TYPE: DNA
ORGANISM: Homo sapiens
US-10-152-724A-24

Query Match 67.9%; Score 22.4; DB 9; Length 36303;
Best Local Similarity 81.2%; Pred. No. 82;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGTTACTTTCTTCTCTATTCAGCC 32
Db 14710 GCTGATGTCACCTACTTCTGTATTCACACC 14741

RESULT 3
US-09-864-761-25312
Sequence 25312, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 25312
LENGTH: 144
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121953.10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EST HUMAN HIT: BE786184.1, EVALUATE 1.00e-20
OTHER INFORMATION: NT HIT: g17706668, EVALUATE 7.00e-21
US-09-864-761-25312

Query Match 65.5%; Score 21.6; DB 10; Length 144;
Best Local Similarity 85.7%; Pred. No. 64;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTCTATTCAG 30
Db 29 TTTGTCTTCTTCTTCTCTCTCTG 56

RESULT 4
US-09-796-692-6086/c
Sequence 6086, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
FILE REFERENCE: HEVATOLOGICAL MALIGNANCIES
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 2001-03-01
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6086
LENGTH: 310
TYPE: DNA

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (184)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (201)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (213)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (217)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (260)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (281)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6086
```

Query Match 65.5%; Score 21.6; DB 9; Length 310;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 3 TGTGTACTTCTCTCTCTATTCAG 30
Db 70 TTTGTCTCTCTCTCTCTCTCTAG 43
```

```
RESULT 5
US-10-040-862-6086/c
Sequence 6086, Application US/10040862
Publication No. US20030078396A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
```

```
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6086
LENGTH: 310
```

```
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (184)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (201)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (213)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (217)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (260)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (281)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-6086
```

Query Match 65.5%; Score 21.6; DB 9; Length 310;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 3 TGTGTACTTCTCTCTCTATTCAG 30
Db 70 TTTGTCTCTCTCTCTCTCTAG 43
```

```
RESULT 6
US-09-796-692-4667/c
Sequence 4667, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Corixa Corporation
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
FILE REFERENCE: 2077,001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
```

```
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4667
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4667

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCATTCGAG 30
    |||||
Db 70 TTTGTTCTTTCTTCTTCCTCTCTAG 43

RESULT 7
US-09-796-692-5376/c
; Sequence 5376, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5376
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
```

```
; LOCATION: (165)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-5376

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCATTCGAG 30
    |||||
Db 70 TTTGTTCTTTCTTCTTCCTCTCTAG 43

RESULT 8
US-09-796-692-7106/c
; Sequence 7106, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7106
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-7106

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCATTCGAG 30
    |||||
Db 70 TTTGTTCTTTCTTCTTCCTCTCTAG 43

RESULT 9
US-10-040-862-4667/c
; Sequence 4667, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
```



```
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
PRIOR FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version:3.0
SEQ ID NO 4667
LENGTH: 408
TYPE: DNA
ORGANISM: Homo sapiens
US-10-040-862-4667

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTTACTTTTCTTCTTCTATTCAG 30
Db 70 TTTGTTCTTTCTTCTTCTTCTCTAG 43

RESULT 10
US-10-040-862-5376/c
; Sequence 5376, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
```

```
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5376
LENGTH: 408
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (165)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-5376

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTTACTTTTCTTCTTCTATTCAG 30
Db 70 TTTGTTCTTTCTTCTTCTTCTCTAG 43

RESULT 11
US-10-040-862-7106/c
; Sequence 7106, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
```

```

; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: PaucSeq for Windows Version 3.0
; SEQ ID NO 7106
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-7106

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTTACTTTCTTCTTCTATCCAG 30
DB 70 TTGTGTTCTTTCTTCTTCTCTCTAG 43

RESULT 12
US-09-783-590-151/C
; Sequence 151, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 151
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (344)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature

```

```

; LOCATION: (407)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (419)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (436)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (439)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (449)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (450)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (457)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (461)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (482)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-151

```

```

Query Match      65.5%; Score 21.6; DB 10; Length 499;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 TGTGTTACTTTCTTCTTCTATCCAG 30
DB 85 TTGTGTTCTTTCTTCTTCTCTCTTAG 58

```

```

RESULT 13
US-09-864-761-8590
; Sequence 8590, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8590
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121953.10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
US-09-864-761-8590

Query Match 65.5%; Score 21.6; DB 10; Length 568;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 97 TTTGTTCCTTCTTCTTCTTCTAG 124

RESULT 14
US-09-822-846-293/C
Sequence 293, Application US/09622846
Publication No. US20030027139A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Pechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakkar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT APPLICATION NUMBER: US/09/822,846
CURRENT FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 293
LENGTH: 1057
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-846-293

Query Match 65.5%; Score 21.6; DB 9; Length 1057;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 549 TTTGTTCCTTCTTCTTCTTCTAG 522

RESULT 15
US-09-938-842A-5185
Sequence 5185, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 5185
LENGTH: 356
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-5185

Query Match 64.8%; Score 21.4; DB 9; Length 356;
Best Local Similarity 80.6%; Pred. No. 87;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAGCA 33
DB 277 TGTGTTCCTTCTTCTTCTTCTATTCAGCA 307

Search completed: May 11, 2003, 06:28:01
Job time : 36.6692 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 328.336 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-17

Perfect score: 31

Sequence: 1 attccacattcattccgcccttcattg 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rtd:*

36: em_hcg_mam:*

37: em_hcg_vtc:*

38: em_sy:*

39: em_hlgo_hum:*

40: em_hlgo_mus:*

41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	100.0	31	6	AX036751	AX036751 Sequence
2	22.6	72.9	152846	2	AC128051	AC128051 Rattus no
3	22	71.0	170897	9	AC020718	AC020718 Homo sapi
4	22	71.0	186665	2	AC101918	AC101918 Mus muscu
5	22	71.0	201646	2	AC116325	AC116325 Mus muscu
6	21.6	69.7	140855	2	AC094340	AC094340 Rattus no
7	21.4	69.0	71489	2	AC124807	AC124807 Mus muscu
8	21.4	69.0	120953	2	AC102902	AC102902 Mus muscu
9	21.4	69.0	194106	2	AC11065	AC11065 Mus muscu
10	21.4	69.0	346897	1	AP002995	AP002995 Mesorhizo
11	21.2	68.4	156159	2	AC023325	AC023325 Homo sapi
12	21.2	68.4	157533	9	AC092868	AC092868 Homo sapi
13	21.2	68.4	189999	9	AC092756	AC092756 Homo sapi
14	21.2	67.7	106755	2	AL138779	AL138779 Homo sapi
15	21	67.7	129121	2	AC120343	AC120343 Mus muscu
16	21	67.7	134011	9	AL354883	AL354883 Human DNA
17	21	67.1	191453	2	AC110136	AC110136 Rattus no
18	20.8	67.1	191453	2	AL840639	AL840639 Mus muscu
19	20.6	66.5	144014	2	AC098695	AC098695 Oryza sat
20	20.6	66.5	159702	9	AL356433	AL356433 Human DNA
21	20.6	66.5	165423	2	AC008111	AC008111 Homo sapi
22	20.6	66.5	167042	9	AC117527	AC117527 Homo sapi
23	20.6	66.5	170216	2	AC012315	AC012315 Homo sapi
24	20.6	66.5	186932	2	AC113061	AC113061 Mus muscu
25	20.6	66.5	233606	2	AL831778	AL831778 Mus muscu
26	20.4	65.8	18201	3	DVDMNPTP	DVDMNPTP
27	20.4	65.8	26435	3	U00034	U00034 Caenorhabdi
28	20.4	65.8	63955	9	AC108151	AC108151 Homo sapi
29	20.4	65.8	78413	2	AC006497	AC006497 Drosophill
30	20.4	65.8	87780	3	AC006714	AC006714 Caenorhab
31	20.4	65.8	92013	2	AC111814	AC111814 Rattus no
32	20.4	65.8	111547	2	AC094583	AC094583 Rattus no
33	20.4	65.8	111967	2	AC010029	AC010029 Drosophill
34	20.4	65.8	124040	2	AC097412	AC097412 Rattus no
35	20.4	65.8	136186	2	AP003896	AP003896 Oryza sat
36	20.4	65.8	141320	2	AC017833	AC017833 Drosophill
37	20.4	65.8	143687	9	AC016572	AC016572 Homo sapi
38	20.4	65.8	147570	2	AC091874	AC091874 Homo sapi
39	20.4	65.8	151514	2	AC113827	AC113827 Rattus no
40	20.4	65.8	153267	2	AC101950	AC101950 Mus muscu
41	20.4	65.8	163191	2	AC107585	AC107585 Rattus no
42	20.4	65.8	169069	2	AC016287	AC016287 Homo sapi
43	20.4	65.8	171612	2	AC068437	AC068437 Homo sapi
44	20.4	65.8	173381	2	AC109087	AC109087 Rattus no
45	20.4	65.8	174805	2	AC107678	AC107678 Mus muscu

ALIGNMENTS

RESULT 1

AX036751

LOCUS

DEFINITION Sequence 17 from Patent WO0058485.

ACCESSION AX036751

VERSION AX036751.1 GI:11226260

KEYWORDS

ORGANISM

SOURCE

synthetic construct.

synthetic construct.

artificial sequence.

1 (bases 1 to 31)

REFERENCE

AUTHORS Rance, I., Theisen, M. and Gruber, V.

TITLE Chimeric expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus

JOURNAL Patent: WO 0058485-A 17 05-OCT-2000;


```

* 34140 34239: gap of unknown length
* 34240 36862: contig of 2623 bp in length
* 36863 36962: gap of unknown length
* 36963 39221: contig of 2259 bp in length
* 39222 39321: gap of unknown length
* 39322 42654: contig of 3333 bp in length
* 42655 42754: gap of unknown length
* 42755 45445: contig of 2691 bp in length
* 45446 45545: gap of unknown length
* 45546 49130: contig of 3585 bp in length
* 49131 49230: gap of unknown length
* 49231 51555: contig of 2325 bp in length
* 51556 51655: gap of unknown length
* 51656 54244: contig of 2589 bp in length
* 54245 54344: gap of unknown length
* 54345 58069: contig of 3725 bp in length
* 58070 58169: gap of unknown length
* 58170 62125: contig of 3956 bp in length
* 62126 62225: gap of unknown length
* 62226 65502: contig of 3277 bp in length
* 65503 65602: gap of unknown length
* 65603 70035: contig of 4434 bp in length
* 70037 70136: gap of unknown length
* 70137 73239: contig of 3103 bp in length
* 73240 73339: gap of unknown length
* 73340 77890: contig of 4551 bp in length
* 77891 77990: gap of unknown length
* 77991 85351: contig of 7361 bp in length
* 85352 85451: gap of unknown length
* 85452 89326: contig of 3875 bp in length
* 89327 89426: gap of unknown length
* 89427 95931: contig of 6505 bp in length
* 95932 96031: gap of unknown length
* 96032 101741: contig of 5710 bp in length
* 101742 101841: gap of unknown length
* 101842 108438: contig of 6597 bp in length
* 108439 108538: gap of unknown length
* 108539 115412: contig of 6874 bp in length
* 115413 115512: gap of unknown length
* 115513 124473: contig of 8961 bp in length
* 124474 124573: gap of unknown length
* 124574 133635: contig of 9061 bp in length
* 133635 133734: gap of unknown length
* 133735 143824: contig of 10090 bp in length
* 143825 143925: gap of unknown length
* 143925 152846: contig of 8922 bp in length.
*
* Location/Qualifiers
*   source          1..152846
*   /organism="Rattus norvegicus"
*   /db_xref="taxon:10116"
*   /clone="CH230-453C4"

```

```

BASE COUNT  43432 a 30075 c 30565 g 44509 t 4265 others
ORIGIN

```

```

Query Match      72.9%; Score 22.6; DB 2; Length 152846;
Best Local Similarity 86.2%; Pred. No. 47;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY  2  TTCCACCTTTTACATTCGCCCTTTCATT 30
Db  36135  TTCCACCTTTCATTCGCCCTTTCCTT 36107

```

```

RESULT 3
LOCUS      AC020718      170897 bp      DNA      linear      PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-472L21 from 2, complete sequence.
ACCESSION  AC020718
VERSION     AC020718.7  GI:15383874
KEYWORDS   HTG.
SOURCE     Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS    Sulston, J.E. and Watson, R.
TITLE      Toward a complete human genome sequence
JOURNAL    Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE    99063792
PUBMED     9847074
REFERENCE
AUTHORS    Cordes, M., Haakena, W. and Bernard, K.
TITLE      The sequence of Homo sapiens BAC clone RP11-472L21
JOURNAL    Unpublished (2001)
REFERENCE
AUTHORS    Waterston, R.H.
TITLE      Direct Submission
JOURNAL    Submitted (08-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE
AUTHORS    Waterston, R.H.
TITLE      Direct Submission
JOURNAL    Submitted (31-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE
AUTHORS    Waterston, R.
TITLE      Direct Submission
JOURNAL    Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            (bases 1 to 170897)
REFERENCE
AUTHORS    Waterston, R.
TITLE      Direct Submission
JOURNAL    Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Aug 31, 2001 this sequence version replaced gi:14030137.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapins@watson.wustl.edu
          Sumary Statistics
Center project name: H_NH0472L21

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPC11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frangen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

AUTHORS

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barina,N., Baetien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 186665)

AUTHORS

Birtten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barina,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,P., Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,J., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17060694.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L17743
Center clone name: 495_1_3

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180378 bases at least Q40
Consensus quality: 183287 bases at least Q30
Consensus quality: 184280 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 184765; sum-of-contigs
Quality coverage: 6.6 in Q20 bases; agarose-fp
Quality coverage: 6.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1081: contig of 1081 bp in length
* 1082 1181: gap of 100 bp
* 1182 2119: contig of 938 bp in length
* 2120 2219: gap of 100 bp
* 2220 3423: contig of 1204 bp in length
* 3424 3523: gap of 100 bp
* 3524 4709: contig of 1186 bp in length
* 4710 4809: gap of 100 bp
* 4810 6162: contig of 1353 bp in length
* 6163 6262: gap of 100 bp
* 6263 7591: contig of 1328 bp in length
* 7592 7691: gap of 100 bp
* 7692 9907: contig of 2216 bp in length
* 9908 10007: gap of 100 bp
* 10008 12103: contig of 2096 bp in length
* 12104 12203: gap of 100 bp
* 12204 16361: contig of 4158 bp in length
* 16362 16461: gap of 100 bp
* 16462 22485: contig of 6024 bp in length
* 22486 22585: gap of 100 bp
* 22586 30730: contig of 8145 bp in length
* 30731 30830: gap of 100 bp
* 30831 40176: contig of 9346 bp in length
* 40177 40276: gap of 100 bp
* 40277 50333: contig of 10117 bp in length
* 50394 50493: gap of 100 bp
* 50494 63206: contig of 12713 bp in length
* 63207 63306: gap of 100 bp
* 63307 77181: contig of 13875 bp in length
* 77182 77281: gap of 100 bp
* 77282 90693: contig of 13412 bp in length
* 90694 90793: gap of 100 bp
* 90794 105614: contig of 14821 bp in length
* 105615 105714: gap of 100 bp
* 105715 122131: contig of 16417 bp in length
* 122132 122231: gap of 100 bp
* 122232 143060: contig of 20829 bp in length
* 143061 143160: gap of 100 bp
* 143161 186665: contig of 43505 bp in length.

FEATURES

source

1. 186665
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1fb="RPCT-24 Male Mouse BAC"
1. 1081
/note="assembly_fragment"
1182. 2119
/note="assembly_fragment"
2220. 3423
/note="assembly_fragment"
3524. 4709
/note="assembly_fragment"
4810. 6162
/note="assembly_fragment"
6263. 7591
/note="assembly_fragment"
7692. 9907
/note="assembly_fragment"
10008. 12103
/note="assembly_fragment"
12204. 16361
/note="assembly_fragment"
16462. 22485
/note="assembly_fragment"
22586. 30730
/note="assembly_fragment"

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0295K03

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET, 0% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 198905 bases at least Q40

```

RESULT	6
AC094340	
LOCUS	AC094340
DEFINITION	Rattus norvegicus clone CH230-3J16, *** SEQUENCING IN PROGRESS ***
	78 unordered pieces.

ACCESSION AC094340
VERSION AC094340.3 GI:21716583
KEYWORDS HTG; HGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Ekaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 140855)
AUTHORS Mizny, D.M., Adams, C., Adio-Ochola, B., Ali-osman, F.R., Allen, C.,

Alshbrook, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, R.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falle, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,
Gabris, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gortell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, U., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulsegged, H.,
Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Monabaci, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwuo, S., Ogun, M., Okunodu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Scherrer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,
Sodergren, E., Sonakke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swatek, A., Tabor, P., Tamerias, A., Tamerias, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 140855)
AUTHORS Worley, K.C.

Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA.

REFERENCE 3 (bases 1 to 140855)
AUTHORS Worley, K.C.

Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941061.

COMMENT

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: GAKM
Center clone name: CH230-3J16
Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 58070 bases at least Q40
Consensus quality: 65587 bases at least Q30
Consensus quality: 71764 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 78 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preceeded.

1 1091: contig of 1091 bp in length
1092 1191: gap of unknown length
1192 2379: contig of 1188 bp in length
2380 2479: gap of unknown length
2480 3753: contig of 1274 bp in length
3754 3853: gap of unknown length
3854 5013: contig of 1160 bp in length
5014 5113: gap of unknown length
5114 6170: contig of 1057 bp in length
6171 6270: gap of unknown length
6271 7273: contig of 1003 bp in length
7274 7373: gap of unknown length
7374 8440: contig of 1067 bp in length
8441 8540: gap of unknown length
8541 9622: contig of 1082 bp in length
9623 9723: gap of unknown length
9723 10887: contig of 1165 bp in length
10888 10987: gap of unknown length
10988 12161: contig of 1174 bp in length
12162 12261: gap of unknown length
12262 13378: contig of 1117 bp in length
13379 13478: gap of unknown length
13479 14576: contig of 1098 bp in length
14577 14676: gap of unknown length
14677 15862: contig of 1186 bp in length
15863 15964: gap of unknown length
15964 16964: contig of 1002 bp in length
16965 17064: gap of unknown length
17065 18127: contig of 1063 bp in length
18128 18227: gap of unknown length
18228 19346: contig of 1119 bp in length
19347 19446: gap of unknown length
19447 20474: contig of 1028 bp in length
20475 20574: gap of unknown length
20575 21588: contig of 1015 bp in length
21590 21689: gap of unknown length
21690 22820: contig of 1131 bp in length
22821 22920: gap of unknown length
22921 23935: contig of 1015 bp in length
23936 24035: gap of unknown length
24036 25086: contig of 1051 bp in length
25087 25186: gap of unknown length
25187 26432: contig of 1246 bp in length
26433 26532: gap of unknown length
26533 27689: contig of 1157 bp in length
27690 27789: gap of unknown length
27790 28808: contig of 1019 bp in length
28809 28908: gap of unknown length
28909 30260: contig of 1352 bp in length
30261 30360: gap of unknown length
30361 31467: contig of 1107 bp in length
31468 31567: gap of unknown length
31568 32637: contig of 1070 bp in length
32638 32737: gap of unknown length
32738 33887: contig of 1150 bp in length
33888 33987: gap of unknown length

```

* 33988 35644: contig of 1657 bp in length
* 35645 35744: gap of unknown length
* 35745 35806: contig of 1062 bp in length
* 36807 36906: gap of unknown length
* 36907 38037: contig of 1131 bp in length
* 38038 38137: gap of unknown length
* 38138 39567: contig of 1430 bp in length
* 39568 39668: gap of unknown length
* 40801 40901: contig of 1134 bp in length
* 40902 40902: gap of unknown length
* 42348 42348: contig of 1447 bp in length
* 42349 42448: gap of unknown length
* 42449 43984: contig of 1536 bp in length
* 43985 44084: gap of unknown length
* 44085 45144: contig of 1060 bp in length
* 45145 45244: gap of unknown length
* 45245 45766: contig of 1522 bp in length
* 46767 46867: gap of unknown length
* 47868 47967: gap of 1001 bp in length
* 47968 49241: contig of 1274 bp in length
* 49242 49342: gap of unknown length
* 49343 50451: contig of 1110 bp in length
* 50452 50551: gap of unknown length
* 50552 52685: contig of 2133 bp in length
* 52685 52784: gap of unknown length
* 52785 54915: contig of 2131 bp in length
* 54916 55015: gap of unknown length
* 55016 56201: contig of 1186 bp in length
* 56202 56301: gap of unknown length
* 56302 57981: contig of 1680 bp in length
* 57982 58081: gap of unknown length
* 58082 59684: contig of 1603 bp in length
* 59685 59784: gap of unknown length
* 59785 61419: contig of 1635 bp in length
* 61420 61519: gap of unknown length
* 61520 62778: contig of 1259 bp in length
* 62779 62878: gap of unknown length
* 62879 63942: contig of 1064 bp in length
* 63943 64042: gap of unknown length
* 64043 65840: contig of 1798 bp in length
* 65841 65940: gap of unknown length
* 65941 67504: contig of 1564 bp in length
* 67505 67604: gap of unknown length
* 67605 69282: contig of 1678 bp in length
* 69283 69383: gap of unknown length
* 69383 70606: contig of 1224 bp in length
* 70607 70706: gap of unknown length
* 70707 73687: contig of 2981 bp in length
* 73688 73787: gap of unknown length
* 73788 76196: contig of 2409 bp in length

```

```

Query Match 69.74; Score 21.6; DB 2; Length 140855;
Best Local Similarity 85.74; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 TCCACCTTTTACATCCGCGCTTTCAT 30
Db 54797 TCCACCTTTTCCCATCCGCGCTTTCAT 54824

```

```

RESULT 7
AC124807/c 71489 bp DNA linear HTG 17-JUN-2002
LOCUS AC124807
DEFINITION Mus musculus clone RP23-137113, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC124807
VERSION AC124807.1 GI:21431243
KEYWORDS HTG; HTGS PHASED.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 71489)

```

AUTHORS JOURNAL TITLE COMMENT

```

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus, clone RP23-137113
TITLE Unpublished
COMMENT 2 (bases 1 to 71489)
Anderson, S., Barta, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazeres, R., Landers, T., Lehoccky, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Themann, N.,
Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Teste, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L26859
Center clone name: 137_I_13
-----
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 758 857: gap of 100 bp
* 858 1599: contig of 742 bp in length
* 1600 1699: gap of 100 bp
* 1700 2430: contig of 731 bp in length
* 2431 2530: gap of 100 bp
* 2531 3268: contig of 738 bp in length
* 3269 3368: gap of 100 bp
* 3369 4117: contig of 749 bp in length
* 4118 4217: gap of 100 bp
* 4218 4978: contig of 761 bp in length
* 4979 5078: gap of 100 bp
* 5079 5820: contig of 742 bp in length
* 5821 5920: gap of 100 bp
* 5921 6676: contig of 756 bp in length
* 6677 6776: gap of 100 bp
* 6777 7540: contig of 764 bp in length
* 7541 7640: gap of 100 bp
* 7641 8387: contig of 747 bp in length
* 8388 8487: gap of 100 bp
* 8488 9229: contig of 742 bp in length

```

```

* 9230 9329: gap of 100 bp
* 9330 10077: contig of 748 bp in length
* 10078 10177: gap of 100 bp
* 10178 10933: contig of 756 bp in length
* 10934 11033: gap of 100 bp
* 11034 11754: contig of 721 bp in length
* 11755 11854: gap of 100 bp
* 11855 12598: contig of 744 bp in length
* 12599 12698: gap of 100 bp
* 12699 13451: contig of 753 bp in length
* 13452 13551: gap of 100 bp
* 13552 14310: contig of 759 bp in length
* 14311 14410: gap of 100 bp
* 14411 15170: contig of 760 bp in length
* 15171 15270: gap of 100 bp
* 15271 16005: contig of 736 bp in length
* 16007 16106: gap of 100 bp
* 16107 16828: contig of 722 bp in length
* 16829 16928: gap of 100 bp
* 16929 17653: contig of 725 bp in length
* 17654 17753: gap of 100 bp
* 17754 18494: contig of 741 bp in length
* 18495 18594: gap of 100 bp
* 18595 19338: contig of 744 bp in length
* 19339 19438: gap of 100 bp
* 19439 20169: contig of 731 bp in length
* 20170 20269: gap of 100 bp
* 20270 21030: contig of 761 bp in length
* 21031 21130: gap of 100 bp
* 21131 21886: contig of 756 bp in length
* 21887 21986: gap of 100 bp
* 21987 22734: contig of 748 bp in length
* 22735 22834: gap of 100 bp
* 22835 23598: contig of 765 bp in length
* 23600 23699: gap of 100 bp
* 23700 24455: contig of 756 bp in length
* 24456 24555: gap of 100 bp
* 24556 25306: contig of 751 bp in length
* 25307 25406: gap of 100 bp
* 25407 25146: contig of 740 bp in length
* 26147 26246: gap of 100 bp
* 26247 26989: contig of 743 bp in length
* 26990 27089: gap of 100 bp
* 27090 27846: contig of 757 bp in length
* 27847 27946: gap of 100 bp
* 27947 28627: contig of 681 bp in length
* 28628 28727: gap of 100 bp
* 28728 29469: contig of 742 bp in length
* 29470 29569: gap of 100 bp
* 29570 30329: contig of 760 bp in length
* 30330 30429: gap of 100 bp
* 30430 31184: contig of 755 bp in length
* 31185 31284: gap of 100 bp
* 31285 32037: contig of 753 bp in length
* 32038 32137: gap of 100 bp
* 32138 32893: contig of 756 bp in length
* 32894 32993: gap of 100 bp
* 32994 33741: contig of 748 bp in length
* 33742 33841: gap of 100 bp
* 33842 34573: contig of 732 bp in length
* 34574 34673: gap of 100 bp
* 34674 35394: contig of 721 bp in length
* 35395 35494: gap of 100 bp
* 35495 36212: contig of 718 bp in length
* 36213 36312: gap of 100 bp
* 36313 37065: contig of 753 bp in length
* 37066 37165: gap of 100 bp
* 37166 37887: contig of 722 bp in length
* 37888 37987: gap of 100 bp
* 37988 38722: contig of 735 bp in length
* 38723 38822: gap of 100 bp
* 38823 39561: contig of 739 bp in length
* 39562 39661: gap of 100 bp

```

```

* 39662 40416: contig of 755 bp in length
* 40417 40516: gap of 100 bp
* 40517 41269: contig of 753 bp in length
* 41270 41369: gap of 100 bp
* 41370 42130: contig of 761 bp in length
* 42131 42230: gap of 100 bp
* 42231 42996: contig of 766 bp in length
* 42997 43096: gap of 100 bp
* 43097 43845: contig of 749 bp in length
* 43846 43945: gap of 100 bp
* 43946 44650: contig of 705 bp in length
* 44651 44750: gap of 100 bp
* 44751 45497: contig of 747 bp in length
* 45498 45597: gap of 100 bp
* 45598 46366: contig of 769 bp in length
* 46367 46466: gap of 100 bp
* 46467 47206: contig of 740 bp in length
* 47207 47306: gap of 100 bp
* 47307 48051: contig of 745 bp in length
* 48052 48151: gap of 100 bp
* 48152 48885: contig of 734 bp in length
* 48886 48985: gap of 100 bp
* 48986 49744: contig of 759 bp in length
* 49745 49844: gap of 100 bp
* 49845 50610: contig of 766 bp in length
* 50611 50710: gap of 100 bp
* 50711 51463: contig of 753 bp in length
* 51464 51563: gap of 100 bp
* 51564 52299: contig of 736 bp in length
* 52300 52399: gap of 100 bp
* 52400 53124: contig of 725 bp in length
* 53125 53224: gap of 100 bp
* 53225 53955: contig of 731 bp in length
* 53956 54055: gap of 100 bp
* 54056 54768: contig of 713 bp in length
* 54769 54868: gap of 100 bp
* 54869 55618: contig of 750 bp in length
* 55619 55718: gap of 100 bp
* 55719 56451: contig of 733 bp in length
* 56452 56551: gap of 100 bp
* 56552 57278: contig of 727 bp in length
* 57279 57378: gap of 100 bp
* 57379 58115: contig of 737 bp in length
* 58116 58215: gap of 100 bp

```

Query Match 69.0%; Score 21.4; DB 2; Length 71489;
 Best Local Similarity 80.6%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 ATTCACCTTTTACATTCGCCCTTTCATTG 31
 Db 28853 ATTTCACATTTTTCATTCCTCTTATATG 28823

RESULT 8 AC102902

LOCUS Mus musculus clone RP24-313G10, WORKING DRAFT SEQUENCE, 4 unordered
 DEFINITION

AC102902 120953 bp DNA linear HTG 20-AUG-2002
 AC102902.2 GI:22325310

ACCESSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 VERSION

KEYWORDS house mouse.
 SOURCE

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 120953)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.
 TITLE Mus musculus, clone RP24-313G10

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 120953)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguski,K., Bouckgalter,B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Glnde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazeres, R., Landers, T., Lepoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McHeeters, R., Meltrin, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talama, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 120953)
AUTHORS
Bjaren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meltrin, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talama, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
On Aug 20, 2002 this sequence version replaced g1:11061012.
COMMENT
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L20056
Center clone name: 313 G.10
Summary Statistics
Sequencing Vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 120411 bases at least Q40
Consensus quality: 120487 bases at least Q30
Consensus quality: 120554 bases at least Q20
Insert size: 115000; agarose-fp
Insert size: 120653; sum-of-ctnigs
Quality coverage: 11.1 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 7151: contig of 7151 bp in length
* 7152 7251: gap of 100 bp
* 7252 15261: contig of 8010 bp in length
* 15262 15361: gap of 100 bp
* 15362 73016: contig of 57655 bp in length
* 73017 73116: gap of 100 bp
* 73117 120953: contig of 47837 bp in length.
Location/Qualifiers
1. 120953
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-313G10"
/clone_lib="RP24-313G10"
1. 7151
/note="assembly_fragment"
7252. 15261
/note="assembly_fragment"
15362. 73016
/note="assembly_fragment"
73117. 120953
/note="assembly_fragment"
clone_end:17
vector_side:right"
BASE COUNT 36211 a 22789 c 22795 g 38857 t 301 others
ORIGIN
Query Match 69.0%; Score 21.4; DB 2; Length 120953;
Best Local Similarity 80.6%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATTCACCTTTTACATTCCTCCCTTCATTCG 31
Db 105772 ATTTAACTTTTACATTCCTCCCTTCATTCG 105802

RESULT 9
AC121085
LOCUS
DEFINITION
Mus musculus clone RP24-494H11, WORKING DRAFT SEQUENCE, 14 ordered pieces.
AC121085
AC121085
VERSION
AC121085.1 GI:20800155
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194106)
Mus musculus, clone RP24-494H11
Unpublished
2 (bases 1 to 194106)

REFERENCE
AUTHORS
Bjaren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Glnde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazeres, R., Landers, T., Lepoczky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meltrin, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,

TITLE
JOURNAL
COMMENT

Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strass, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L21146

Center clone name: 494_H_11

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 189377 bases at least Q40

Consensus quality: 191587 bases at least Q30

Consensus quality: 192278 bases at least Q20

Insert size: 18800; agarose-fp

Quality coverage: 6.0 in Q20 bases; agarose-fp

Quality coverage: 5.8 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```
1 1278: contig of 1278 bp in length
1279 1378: gap of 100 bp
1379 2686: contig of 1308 bp in length
2687 2786: gap of 100 bp
2787 4430: contig of 1644 bp in length
4431 4530: gap of 100 bp
4531 8486: contig of 3956 bp in length
8487 8586: gap of 100 bp
8587 12856: contig of 4270 bp in length
12857 12956: gap of 100 bp
12957 22063: contig of 9107 bp in length
22064 22163: gap of 100 bp
22164 30661: contig of 8498 bp in length
30662 30761: gap of 100 bp
30762 42347: contig of 11586 bp in length
42348 42447: gap of 100 bp
42448 55768: contig of 13321 bp in length
55769 55868: gap of 100 bp
55869 77586: contig of 21718 bp in length
77587 77686: gap of 100 bp
77687 103032: contig of 25346 bp in length
103033 103132: gap of 100 bp
103133 127989: contig of 24857 bp in length
127990 128089: gap of 100 bp
128090 155766: contig of 27679 bp in length
155769 155868: gap of 100 bp
155869 194106: contig of 38238 bp in length.
```

FEATURES

source

```
1. 194106
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-494H11"
/clone_1b="RPC1-24 Male Mouse BAC"
```

```
misc_feature 1. 1278
/notes="assembly_fragment"
misc_feature 1379. 2686
/notes="assembly_fragment"
misc_feature 2787. 4430
/notes="assembly_fragment"
misc_feature 4531. 8486
/notes="assembly_fragment"
misc_feature 8587. 12856
/notes="assembly_fragment"
misc_feature 12957. 22063
/notes="assembly_fragment"
misc_feature 22164. 30661
/notes="assembly_fragment"
misc_feature 30762. 42347
/notes="assembly_fragment"
misc_feature 42448. 55768
/notes="assembly_fragment"
misc_feature 55869. 77586
/notes="assembly_fragment"
misc_feature 77687. 103032
/notes="assembly_fragment"
misc_feature 103133. 127989
/notes="assembly_fragment"
misc_feature 128090. 155768
/notes="assembly_fragment"
misc_feature 155869. 194106
/notes="assembly_fragment"
```

BASE COUNT 54107 a 42575 c 42309 g 53814 t 1301 others

ORIGIN

Query Match 69.0%; Score 21.4; DB 2; Length 194106;

Best local Similarity 80.6%; Pred. No. 1.5e+02;

Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATTCACCTTTTACATTCGCCCTTCATTG 31

Db 171991 ATTCACCTCTTACGTTTCATCTTTCATTG 172021

```
RESULT 10
AP002995/c 346897 bp DNA linear BCT 15-MAY-2001
LOCUS Mesorhizobium loti DNA, complete genome, section 2/21.
DEFINITION AP002995 BA000012
ACCESSION AP002995.2 GI:114021442
VERSION
KEYWORDS
SOURCE
ORGANISM Mesorhizobium loti (strain:MAFP30309) DNA.
```

REFERENCE 1 (sites)
Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.

AUTHORS

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti
Bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)

JOURNAL 2 (bases 1 to 346897)

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp;
<http://www.kazusa.or.jp/rhizobase/>;
Tel: 81-438-52-3935 (ex. 2338); Fax: 81-438-52-3934)

On May 11, 2001 this sequence version replaced gi:11994963.
Location/Qualifiers

FEATURES
source

1. 346897

gene /organism="Mesorhizobium loti"
/strain="MAF30309"
/db_xref="taxon:381"
74..160
/note="trnS-TGA"
complement (240..1109)
/gene="ms18587"
complement (240..1109)
/gene="ms18587"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48055.1"
/db_xref="GI:14021443"
/translation="MNDOTOTGTGTTALPYTVVTRBAGDVANPOAVHKGTGPFYHPRG
AGNVPILVSNKHLVCKMLLEPFDALLDNGKRYFGPANTFRLTKOQLPIFEHPDPTV
DLAIP1QPIVEANRGQKKEI FNLLGRETRTAPDNTQAILHAITSVLMVFPESGIDME
KNMLPVVRGILAVNRYLTDYRGETNLAARFGSSGSPVAF FENMLPDGSGGYTM
LAOPVYFLVLSHSGPMTATGKIVPVPTSDQIAOTQMLHLGYCVKASRYELVE
TIRAKISEPTASG"
complement (1118..1657)
/gene="m110476"
complement (1118..1657)
/gene="m110476"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48056.1"
/db_xref="GI:14021444"
/translation="MFPISALIGFEKDGQWRGALALAGEAMAQTAEENPFIEMVREE
KNLVERICRSPDSFISQAAMAIRRRPGMVLIIHNSRHVMEKIIITGEBKVPQITV
DGSVHAGDVADLGDIREMHLRAMCKNSHAEVKPALLIKRYGEALFSTVERALFC
TGCDDRGSPVLEIHKLP RN"
2193..2558
/gene="m1r0478"
2193..2558
/gene="m1r0478"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48057.1"
/db_xref="GI:14021445"
/translation="MALALGALMKRLLYFVMTLAVTSGCAADYLNMYDTVLASGD
AMTNSLLOTVDPENPNNTHTIEGDQRLIVGVVQRYRSPISAGGYSNGCPDINSVA
ADSHGCKRSABSRPGAM"
2651..3412
/gene="m1r0479"
2651..3412
/gene="m1r0479"
/note="unknown protein"
/codon_start=1
/transl_table=1
/product="sucinoglycan biosynthesis protein; Exo1"
/protein_id="BAB48058.1"
/db_xref="GI:14021446"
/translation="MPLRLVLTVAASALATOPVMNNGSLPEINLONLTKRPVNI
SGVASVIDGPTIEHMGORVFNGLDAPESAQVCDDAKGFYPCGRSABALDGLIAS
KPVQAFVMTDRGRFYGDCTRADGSSVAAVMYEHGOALDMPYXNSGAVAQAKAA
AKGLMIGSFQAPDRRAQSHDDAQTPTAPLPATGNSPCNITKGNISASGERIYHLP
GQRYAVATVISQNKGRMFCSEAEVAAGRRSRK"
3429..3782
/gene="m1r0480"
3429..3782
/gene="m1r0480"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48059.1"
/db_xref="GI:14021447"
/translation="MKRHVRAAVALIASHSSGNNVSYSELTAKVNI DASVQGV
VDAYDNNCHIDGTPSLYHYGNSHTIDLPQSGGQYDGDIGNSHFLVTSNSNA
ELVDYGGGFFSPST"
complement (5516..5947)
/gene="m110481"

CDS complement (5516..5947)
/gene="m110481"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48060.1"
/db_xref="GI:14021448"
/translation="MNSDRGTGKDPSTGRFLPGHSGFKYVGRSRVROAIDTANSHA
ATPMMLVGMKMLAGDIRAEILLRYGADPEFLPKRNVLELDVENPTISEMLATVLSAA
WSGEISPDPAKQLDIAIAKGKIEBAELLRRVEQLAEAYKR"
complement (6065..6883)
/gene="m110482"
complement (6065..6883)
/gene="m110482"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48061.1"
/db_xref="GI:14021449"
/translation="MIDPKRIDGASSRGLSPQMNYSVISAFFMPEYSTMSPDIDA
ASATLNGVLAAGNAVIGAERLKRDETRDVSLEAIARYVAAPRYLADRAIDERN
VGISRNSLAQPIQINPTPTSTAASLEERQVRNMLDGDQIGEAHDMQSGCITVL
NAVAPSIDLTIPSQQLAGLIVEEFAIKRMARTIDNPSNDNLATANPLKNMSDDL
VRLAKARLDGKLREQAVSDAEKLLAVIARASQAGNLTPDEAFKMLGQKVA"
complement (7133..7753)
/gene="m110483"
complement (7133..7753)
/gene="m110483"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48062.1"
/db_xref="GI:14021450"
/translation="MANYITLRLTVMKRASITTDHANPAKLGQIADAPFLMSRT
REWASVNPSEFOYRDECDVWSGEGNIRITFSFKTYQSIFMDAATTDGHPV
WGLTGCCADNMVVSSTHFSFGIINGVSASVQSDMWQDESSSTVSSMLDRLHPW
RDPPTGVAROSKELPPAIRLVKLGYSRSDTRMTWAPNQAFENIH"
complement (7827..10169)
/gene="m110485"
complement (7827..10169)
/gene="m110485"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48063.1"
/db_xref="GI:14021451"
/translation="MRDSSVRLPELLPYRQFVWRLLDQKQSSPKPAMPYSANGCR
ADTNPATWSDYPTAVRALASQYAGIGVFENDPFPVIDLGGCDATGTGLSAAEA
AITRMPNNAWELSGSEGIIHIFGRCDKVALFRKMAWCGCFYTRKRFVAFPGGI
OGNIDIDNTNOLAIIIVARGVGEODASQAVEDPRMIGPDDLELRLFLASTGCTK
AKIOASPTNLOLMEHAALAOVFPBATKPFDDHSAADQALNVLAFTMGRODYAIR
LRRSALGRBKRTRAAYRNTVNAVAVSCQTVYGCNQGRSRROYQENVRIGSDIG
ADITHVWTLIEDLQDLWFIHRSKSGAIAISRTLVLPATDQREYSASREVAAST
IDPKTDGFKRKLVPFNWQSPRSKSVVITWKPNGAVICEAPESQAGFTWRGFO
PYVLPQSOLVLEWSEHLYLPIEAERVREOMLAHIVQPEELPOGAYLFTTET
GIGRPMGOSLVLRVMSHAYLIDIDEMLDGTFNGLSKRLAVVDEAAGQGYGRW
SHSEKLSMINIEMPRINKEHGLEFVEMNCBWLPSNNMDLPPDASRRRVVNAV
THRQSPDYTRITISQIANPDTISAKRLEBLDISPFGKALPDNBAERKALGNAAS
DLNLLAARDEMPGLARSDVINYLHVSIAKLPEKTKINKLIEQAKMLVERISIG
PNKYRVVVDYTAQOIEITNPQWAAVAAVNARQKGFROG"
complement (10162..10392)
/gene="m110486"
complement (10162..10392)
/gene="m110486"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB48064.1"
/db_xref="GI:14021452"
/translation="MECVAMGKLVREMNQSAIYVSIPOVCATTLGISRPTLDRIYKD
NPSFPRKKIIGPLVGLFRSDILETYIGSLPDA"
complement (10595..11803)
/gene="m110487"

gene

CDS
 complement (10595..11803)
 /gene="ml10487"
 /codon_start=1
 /transl_table=11
 /product="probable prophage integrase"
 /protein_id="BAB48065.1"
 /db_xref="GI:14021453"
 /translation="MLTIREIOSLVHKGEPGWSHTKALYFPIASGPRPMIARYMLK
 GRRRSLSDISIPATVRELKQIELSETEBKVRKGIIDPLAEBAKAASPEPVROIV
 TRKSVLDVIAANRGKGNKAKHYOQMENTATVYPVIGLSEPOSLITTEVLKLOPL
 MATKRTASRVSRSVETITGAKAKGSEKPEIWMANHNANRANWADLWFLNKRQEN
 HFSAMPPEIIPAFQSIATERDYSTALQPLILVARTNNTLGAENSEIDTEAVVIT
 PASRMKSGKEHVPVLSSEALKVASLIRIGNPVLPFGALRNLSLNMALMVMRGK
 LGHYVPHGFRSAFRDMAAEETVYPNHIEMALAHITIKNKVEAAYRRGDLMKRRQIME
 DMSAYLASAR"
 12686..13279
 /gene="mlr0488"
 12686..13279
 /gene="mlr0488"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB48066.1"
 69.0%; Score 21.4; DB 1; Length 346897;
 Best Local Similarity 80.6%; Pred. No. 1.5e+02;
 Matches 45; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 ATTCACCTTTACATTCGCCCTTTCATTG 31
 |||||
 Db 37212 ATCCACGCTTTGCTTCACGCTTTCATTG 37182
 |||||

RESULT 11
 AC023325
 LOCUS
 DEFINITION Homo sapiens chromosome 15 clone RP11-356M20 map 15, WORKING DRAFT
 AC023325
 SEQUENCE 15 unordered pieces.
 AC023325 3 GI:12658138
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 156159)
 Birren,B., Linton,L., Nuebaum,C. and Lander,E.
 Homo sapiens chromosome 15, clone RP11-356M20
 Unpublished
 2 (bases 1 to 156159)
 Birren,B., Linton,L., Nuebaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bede,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearlano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
 Feneclor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
 Galasiano,J., Gardyna,S., Ginde,S., Goeyete,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatae,A.,
 Klein,J., Landers,T., Largocque,K., Lehocsky,J., Levine,R.,
 Lileu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McCarthy,M.,
 McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
 Menais,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
 Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Subramanian,A., Talamas,J., Testave,S., Theodore,J., Tirrell,A.,
 Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
 Zody,M.
 Direct Submission
 Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 3, 2001 this sequence version replaced gi:7139826.
 All repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WMR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 356.M.20
 Center clone name: 356.M.20
 ----- Summary Statistics
 Sequencing vector: M13; M77815, 71% of reads
 Sequencing vector: Plasmid; n/a; 29% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 149772 bases at least Q40
 Consensus quality: 152333 bases at least Q30
 Consensus quality: 153705 bases at least Q20
 Insert size: 15700; agarose-fp
 Insert size: 154759; sum-of-contigs
 Quality coverage: 6.4 in Q20 bases; agarose-fp
 Quality coverage: 6.5 in Q20 b.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 9882: contig of 9882 bp in length
 9883 9982: gap of 100 bp
 9983 11301: contig of 1319 bp in length
 11302 11401: gap of 100 bp
 11402 12615: contig of 1214 bp in length
 12616 12715: gap of 100 bp
 12716 16302: contig of 3587 bp in length
 16303 16402: gap of 100 bp
 16403 18619: contig of 2217 bp in length
 18620 18719: gap of 100 bp
 18720 21764: contig of 3045 bp in length
 21765 21864: gap of 100 bp
 21865 66492: contig of 44628 bp in length
 66493 66592: gap of 100 bp
 66593 69609: contig of 3017 bp in length
 69610 69709: gap of 100 bp
 69710 75994: contig of 6285 bp in length
 75995 76094: gap of 100 bp
 76095 82300: contig of 6206 bp in length
 82301 82400: gap of 100 bp
 82401 91215: contig of 8815 bp in length
 91216 91315: gap of 100 bp
 91316 101994: contig of 10679 bp in length
 101995 102094: gap of 100 bp
 102095 128651: contig of 26557 bp in length
 128652 128751: gap of 100 bp
 128752 153403: contig of 24652 bp in length
 153404 153503: gap of 100 bp
 153504 156159: contig of 2656 bp in length.
 Location/Qualifiers
 1..156159
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="RP11-356M20"
 /clone_lib="RP11-11 Human Male BAC"
 1..9882
 /note="assembly_fragment
 clone_end:SP6

```

misc_feature      vector:side:left
                  9983..11301
/note="assembly_fragment"
misc_feature      11402..12615
/note="assembly_fragment"
misc_feature      12716..16302
/note="assembly_fragment"
misc_feature      16403..18619
/note="assembly_fragment"
misc_feature      18720..21764
/note="assembly_fragment"
misc_feature      21865..66492
/note="assembly_fragment"
misc_feature      66593..69609
/note="assembly_fragment"
misc_feature      69710..75994
/note="assembly_fragment"
misc_feature      76095..82300
/note="assembly_fragment"
misc_feature      82401..91215
/note="assembly_fragment"
misc_feature      91316..101994
/note="assembly_fragment"
misc_feature      102095..128651
/note="assembly_fragment"
misc_feature      128752..153403
/note="assembly_fragment"
misc_feature      153504..156159
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
                  clone_end:T7
                  vector:side:right"
BASE COUNT      44142 a 33879 c 33433 g 43302 t 1403 others
ORIGIN
Query Match      68.4%; Score 21.2; DB 2; Length 156159;
Best Local Similarity 88.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 ATTCACCTTTACATTCGCCCTTT 26
Db 72877 ATTCACCTTTACATTCGCCATT 72902

RESULT 12
AC092868      157533 bp DNA linear PRI 22-NOV-2001
LOCUS      Homo sapiens chromosome 15 clone RP11-356M20 map 15q21.3, complete
sequence.
ACCESSION      AC092868
VERSION      AC092868.2 GI:17047069
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 157533)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dore,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Sequencing of human chromosome 15 DIS146-DIS117 region
Unpublished
2 (bases 1 to 157533)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dore,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submision
Submitted (01-AUG-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
3 (bases 1 to 157533)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dore,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

TITLE      Direct Submision
JOURNAL      Submitted (22-NOV-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT      On Nov 22, 2001 this sequence version replaced gi:15055287.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbr.washington.edu/msg_www
Contact: leerowen@systembiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC092756 [drafting center:
UWMSC], and AC091748 [drafting center: UWMSC] were added for
finishing
-----
FEATURES
source      Location/Qualifiers
            1..157533
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15q21.3"
            /clone="RP11-356M20"
            /clone_1lb="RPC1 human BAC library 11"
            /note="Data from overlapping BACs RP11-429D19 and
RP11-99L18 were added and the consensus sequence was
determined from RP11-356M20 to the extent possible"
            1..24730
            /note="overlap with RP11-429D19 AC092756"
            28322..157533
            /note="overlap with RP11-99L18 AC091748"
            147556..147559
            /note="low quality data"
BASE COUNT      45312 a 34917 c 33839 g 43465 t
ORIGIN
Query Match      68.4%; Score 21.2; DB 9; Length 157533;
Best Local Similarity 88.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 ATTCACCTTTACATTCGCCCTTT 26
Db 13088 ATTCACCTTTACATTCGCCATT 13113

RESULT 13
AC092756      189999 bp DNA linear PRI 19-JAN-2002
LOCUS      Homo sapiens chromosome 15 clone RP11-429D19 map 15q21.3, complete
sequence.
ACCESSION      AC092756
VERSION      AC092756.2 GI:18249990
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 189999)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dore,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Sequencing of human chromosome 15 DIS146-DIS117 region
Unpublished
2 (bases 1 to 189999)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dore,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submision
Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute

```

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
 3 (bases 1 to 189999)
 Rowen, L., Madan, A., Qin, S., Baradaran, L., Birditt, B., Bloom, S., Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D., and Hood, L.
 Direct Submission
 Submitted (19-JAN-2002) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
 On Jan 19, 2002 this sequence version replaced gi:15021994.

----- Genome Center
 Center: Multimegabase Sequencing Center
 Center code: UMSC
 Web site: http://chroma.mbc.washington.edu/mag_www
 Contact: leerowen@systembiology.org

----- Summary Statistics
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 90% of reads
 Chemistry: Dye-primer Big Dye; 10% of reads
 Assembly program: Phrap; version 0.990399

Data from overlapping BACs AC092757 [drafting center: UMSC], AC023890 [drafting center: WUGSC], and AC092868 [drafting center: WIBR] was added for finishing

----- Location/Qualifiers
 source 1.189999
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q21.3"
 /clone="RP11-429D19"
 /clone_lib="RPCT human BAC library 11"
 /note="Data from overlapping BACs RP11-59H7, RP11-756K9, and RP11-356M20 was added and the consensus sequence was determined from RP11-429D19 to the extent possible"
 1.147389
 /note="overlap with: RP11-756K9 AC023890"
 1.56595
 /note="overlap with RP11-59H7 AC092757"
 137567
 /note="low quality data"
 164999
 /note="low quality data"
 165270.189999
 /note="overlap with RP11-356M20"
 misc_feature
 BASE COUNT 54028 a 41117 c 41361 g 53493 t
 ORIGIN

Query Match 68.4%; Score 21.2; DB 9; Length 189999;
 Best Local Similarity 88.5%; Pred. No. 1.8e+02;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTCACCTTTACATCCGCCCTT 26
 |||||
 Db 178357 ATTCACCTTTACATTCGCACT 178382

RESULT 14
 AL138779/c 106755 bp DNA linear HTG 10-JUN-2001
 LOCUS Homo sapiens chromosome 1 clone RP6-65P20 map p32.2-34.1, ***
 DEFINITION SEQUENCING IN PROGRESS ***, 20 unordered pieces.
 ACCESSION AL138779
 VERSION AL138779.7 GI:9930811
 KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 106755)
 AUTHORS McIay, K.

TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hunquetry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Aug 27, 2000 this sequence version replaced gi:9926429.

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: hunquetry@sanger.ac.uk

----- Project Information
 Center project name: dA65F20
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 14% of reads
 Chemistry: Dye-terminator ET-amersham; 85% of reads
 Consensus quality: 99905 bases at least Q40
 Consensus quality: 102097 bases at least Q20
 Insert size: 10485; sum-of-contigs
 Insert size: 117633; 9.0% error; agarose-fp
 Quality coverage: 3.06x in Q20 bases; sum-of-contigs Quality coverage: 3.01x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2182: contig of 2182 bp in length
 2183 2282: gap of 100 bp
 2283 11274: contig of 8992 bp in length
 11275 11374: gap of 100 bp
 11375 17141: contig of 5767 bp in length
 17142 17241: gap of 100 bp
 17242 21606: contig of 4365 bp in length
 21607 21706: gap of 100 bp
 21707 26686: contig of 4980 bp in length
 26687 26786: gap of 100 bp
 26787 31996: contig of 5210 bp in length
 31997 32096: gap of 100 bp
 32097 35605: contig of 3503 bp in length
 35606 35705: gap of 100 bp
 35706 49561: contig of 13856 bp in length
 49562 49661: gap of 100 bp
 49662 52981: contig of 3320 bp in length
 52982 53081: gap of 100 bp
 53082 62347: contig of 9266 bp in length
 62348 62447: gap of 100 bp
 62448 66552: contig of 4105 bp in length
 66553 66652: gap of 100 bp
 66653 69451: contig of 2799 bp in length
 69452 69551: gap of 100 bp
 69552 73006: contig of 3455 bp in length
 73007 73106: gap of 100 bp
 73107 77114: contig of 4008 bp in length
 77115 77214: gap of 100 bp
 77215 80477: contig of 3263 bp in length
 80478 80577: gap of 100 bp
 80578 85312: contig of 4735 bp in length
 85313 85412: gap of 100 bp
 85413 89155: contig of 3743 bp in length
 89156 89255: gap of 100 bp
 89256 92958: contig of 3703 bp in length
 92959 93058: gap of 100 bp
 93059 98962: contig of 5904 bp in length
 98963 99062: gap of 100 bp
 99063 106755: contig of 7693 bp in length.

Location/Qualifiers

```

source
1. .106755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p32.2-34.1"
/clone="RP6-65F20"
/clone_id="RP6-65F20"
1. .2182
/note="assembly fragment:00605
fragment chain:1"
clone_end:SP6
vector_side:left"
2283. .11274
/note="assembly fragment:00337
fragment chain:1"
11375. .17141
/note="assembly fragment:00999
fragment chain:2"
17242. .21606
/note="assembly fragment:00285
fragment chain:2"
21707. .26686
/note="assembly fragment:00197
fragment chain:2"
26787. .31996
/note="assembly fragment:00295
fragment chain:2"
32097. .35605
/note="assembly fragment:00488
fragment chain:3"
35706. .49561
/note="assembly fragment:00322
fragment chain:3"
49662. .52981
/note="assembly fragment:00261
fragment chain:3"
53082. .62347
/note="assembly fragment:00320
fragment chain:4"
62448. .66552
/note="assembly fragment:00882
fragment chain:4"
66653. .69451
/note="assembly fragment:00943
fragment chain:5"
69552. .73006
/note="assembly fragment:00204
fragment chain:5"
73107. .77114
/note="assembly fragment:00302"
77215. .80477
/note="assembly fragment:00594"
80578. .85312
/note="assembly fragment:00828"
85413. .89155
/note="assembly fragment:00646
fragment chain:6"
89256. .92958
/note="assembly fragment:00058
fragment chain:6"
93059. .98962
/note="assembly fragment:00457
fragment chain:6"
99063. .106755
/note="assembly fragment:00019
fragment chain:6"
clone_end:T7
vector_side:right"

BASE COUNT 30905 a 21750 c 21309 g 30884 t 1907 others
Origin
Query Match 67.7%; Score 21; DB 2; Length 106755;
Best Local Similarity 82.8%; Pred. No. 2.2e+02;

```

```

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATTGACCTTTACATCCGCCCTTCAT 29
Db 52888 AATTGACCTTTGACATTCCTGCTTCAT 52860

RESULT 15
AC120343/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITL
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC120343 129121 bp DNA linear HTG 21-AUG-2002
Mus musculus clone RP23-149H21, WORKING DRAFT SEQUENCE, 10 ordered
pieces.
AC120343 GI:22381827
HTG; HTGS_PHASR2; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 129121).
Birren, B., Nuebaum, C. and Lander, E.
Mus musculus, clone RP23-149H21
Unpublished
2 (bases 1 to 129121)
Anderson, S., Barna, N., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Birren, B., Linton, L., Bastien, V., Bloom, T., Boguslavsky, L.,
McCarthy, M., McGowan, P., McKernan, K., Meldrum, J., Menue, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 129121)
Birren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Haeos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karacas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, D., Matthews, C.,
McCarthy, M., Meldrum, J., Menue, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:20503130.
All repeats were identified using RepeatMasker:

```

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19845
 Center clone name: 14g H.21
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 126860 bases at least Q40
 Consensus quality: 127509 bases at least Q30
 Consensus quality: 127930 bases at least Q20
 Insert size: 18200; agarose-fp
 Insert size: 18221; sum-of-contigs
 Quality coverage: 7.2 in Q20 bases; agarose-fp
 Quality coverage: 10.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 1147: contig of 1147 bp in length
 * 1148 1247: gap of 100 bp
 * 1248 2769: contig of 1522 bp in length
 * 2770 2869: gap of 100 bp
 * 2870 5607: contig of 2738 bp in length
 * 5608 5707: gap of 100 bp
 * 5708 10061: contig of 4354 bp in length
 * 10062 10161: gap of 100 bp
 * 10162 17248: contig of 7087 bp in length
 * 17249 17348: gap of 100 bp
 * 17349 27373: contig of 10025 bp in length
 * 27374 27473: gap of 100 bp
 * 27474 41686: contig of 14213 bp in length
 * 41687 41786: gap of 100 bp
 * 41787 64402: contig of 22616 bp in length
 * 64403 64502: gap of 100 bp
 * 64503 104541: contig of 40039 bp in length
 * 104542 104641: gap of 100 bp
 * 104642 129121: contig of 24480 bp in length.
 Location/Qualifiers
 1. 129121
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RPC1-23 Female Mouse BAC"
 1. 1147
 /note="assembly_fragment"
 1248. 2769
 /note="assembly_fragment"
 2870. 5607
 /note="assembly_fragment"
 5708. 10061
 /note="assembly_fragment"
 10162. 17248
 /note="assembly_fragment"
 17349. 27373
 /note="assembly_fragment"
 27474. 41686
 /note="assembly_fragment"
 41787. 64402
 /note="assembly_fragment"
 64503. 104541
 misc_feature

misc_feature /note="assembly_fragment"
 104642. 129121
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"
 BASE COUNT 34269 a 28590 c 29000 g 36358 t 904 others
 ORIGIN
 Query Match 67.7%; Score 21; DB 2; Length 129121;
 Best Local Similarity 82.8%; Pred. No. 2.2e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 3 TCCACCTTTACATTCGCCCTTTCATG 31.
 |||||
 Db 46410 TCCACCTTTTCCTTCCACCTTTCATG 46382
 Search completed: May 11, 2003, 06:16:37
 Job time : 570.336 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 530.481 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33
Sequence: 1 gctgtgtactctctctctctctcagccca 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthm:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
C 1	23.6	71.5	996	AL068002 Drosophila
C 2	23.6	71.5	1239	BC258346 602379796
C 3	23.2	70.3	619	BH085145 BH085145
C 4	23	69.7	479	BF397893 UI-R-B52
C 5	23	69.7	1522	AK005882 Mus muscu
C 6	22.6	68.5	634	BQ247773 TAE250426

C 7	22.6	68.5	697	14	BQ249531	BQ249531 TAE250426
C 8	22.6	68.5	1527	12	BE876684	BE876684 601488248
C 9	22.4	67.9	693	17	BH596639	BH596639 BOME9197
C 10	22.2	67.3	455	17	A0872730	A0872730 V15F1 mTh
C 11	22.2	67.3	473	17	A0874006	A0874006 V9G9 mTh
C 12	22.2	67.3	577	17	A0872889	A0872889 V54C5 mTh
C 13	22.2	67.3	708	17	A0872952	A0872952 V56B3 mTh
C 14	22.2	67.3	708	12	BG580976	BG580976 102402980
C 15	22.2	67.3	810	17	A0876136	A0876136 V133F7 mT
C 16	22.2	67.3	887	17	A2685547	A2685547 ENTRE48TR
C 17	22.2	66.7	545	13	BM270934	BM270934 sak03C12
C 18	21.8	66.1	667	17	A2494068	A2494068 1M0339N06
C 19	21.8	66.1	854	12	BF688017	BF688017 602067220
C 20	21.8	66.1	1176	13	BI416154	BI416154 602989006
C 21	21.6	65.5	144	10	BE152640	BE152640 CM1-HT032
C 22	21.6	65.5	375	13	BT007043	BT007043 RCS-RT005
C 23	21.6	65.5	429	11	BC008693	BC008693 Homo_BAP1
C 24	21.6	65.5	443	10	BB847572	BB847572 BB847572
C 25	21.6	65.5	460	10	AW984336	AW984336 PM3-HN001
C 26	21.6	65.5	464	12	BF706719	BF706719 281344 MA
C 27	21.6	65.5	473	10	AW984414	AW984414 PM3-HN001
C 28	21.6	65.5	499	12	BF439546	BF439546 nab67G02
C 29	21.6	65.5	505	9	AI963143	AI963143 wt27H04.X
C 30	21.6	65.5	509	13	BM511124	BM511124 IJ4C10.Y
C 31	21.6	65.5	534	9	AI747788	AI747788 u102d07.Y
C 32	21.6	65.5	536	12	BC986015	BC986015 359762 MA
C 33	21.6	65.5	538	12	BF042283	BF042283 BP250003A
C 34	21.6	65.5	548	12	BE722524	BE722524 190803 MA
C 35	21.6	65.5	563	12	BE835234	BE835234 RCS-FN002
C 36	21.6	65.5	576	12	BG716076	BG716076 602677503
C 37	21.6	65.5	578	17	AQ309992	AQ309992 CITR1-E1
C 38	21.6	65.5	585	14	BQ018935	BQ018935 UI-H-DH1
C 39	21.6	65.5	606	14	BM761228	BM761228 K-EST0041
C 40	21.6	65.5	644	13	BM510872	BM510872 IJ44C10.X
C 41	21.6	65.5	653	12	BE885377	BE885377 601507456
C 42	21.6	65.5	679	12	BG504516	BG504516 602552488
C 43	21.6	65.5	686	14	BQ019003	BQ019003 UI-H-DH1
C 44	21.6	65.5	688	13	BM551574	BM551574 AGENCOURT
C 45	21.6	65.5	691	14	BM677596	BM677596 UI-E-E01

ALIGNMENTS

RESULT 1
CNS00LES/c 996 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC:
BACR30A08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Oono and
Aaron Mammeter in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1..996

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR30A08"

/clone_id="RP1-98"

/note="end : 17"

BASE COUNT 387 a 161 c 149 g 222 t 77 others

ORIGIN

Query Match 71.5%; Score 23.6; DB 17; Length 996;

Best Local Similarity 81.2%; Pred. No. 3.9e+03;

Matches 26; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 GCTGTTACTTTCTCTCTATTCAGCC 32

Db 836 GCTCTTTATTTCTCTCTCTGCGAGCC 805

RESULT 2 BG258346 1239 bp mRNA linear EST 13-FEB-2001

LOCUS 622379796F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510703 5,

DEFINITION mRNA sequence.

ACCESSION BG258346

VERSION BG258346.1 GI:12768162

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 1239)

NIH-MGC <http://mgs.nci.nih.gov/>.

Unpublished (1999)

Contact: Robert Straubeberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: LLM10392 row: c column: 24

High quality sequence stop: 423.

Location/Qualifiers

1..1239

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4510703"

/clone_id="NIH_MGC_92"

/tissue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC library."

BASE COUNT 364 a 366 c 346 g 163 t

ORIGIN

Query Match 71.5%; Score 23.6; DB 12; Length 1239;

Best Local Similarity 86.7%; Pred. No. 3.6e+03;

Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 TGTGTTACTTTCTCTCTATTCAGCC 32

Db 835 TGTGTTACTTTCTCTCTGCGAGCC 806

RESULT 3 BH085145 619 bp DNA linear GSS 18-JUL-2001

LOCUS RPI-24-400H14.TV RPI-24 Mus musculus genomic clone RPI-24-400H14

DEFINITION DNA sequence.

ACCESSION BH085145

VERSION BH085145.1 GI:14904742

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 619)

Zhao, S., Mierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,

Tegayle, G., Geer, K., Krol, M., Shvartsbeyn, A., Gabregiorge, E.,

Russell, D., de Jong, P., and Fraser, C.M.

Mouse BAC End Sequences from Library RPI-24

Unpublished (1999)

Other_GSSs: RPI-24-400H14.TV

Contact: Shaoying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdjejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html

Plate: 400 row: H column: 14

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..619

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPI-24-400H14"

/clone_id="RPI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pPARBAC1; Site 1: BamHI; Site 2: BamHI;

RPI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pPARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 203 a 100 c 124 g 192 t

ORIGIN

Query Match 70.3%; Score 23.2; DB 17; Length 619;

Best Local Similarity 89.3%; Pred. No. 5.6e+03;

Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 CTGTTGTTACTTTCTCTCTATTCACA 29

Db 515 CTGTTGTTACTTTCTCTCTATTCACA 488

RESULT 4 BF397893 479 bp mRNA linear EST 27-NOV-2000

LOCUS UI-R-BS2-bec-h-04-0-UI_a1 UI-R-BS2 Ratius norvegicus cDNA clone

DEFINITION UI-R-BS2-bec-h-04-0-UI 3', mRNA sequence.

ACCESSION BF397893

VERSION BF397893.1 GI:11382878

KEYWORDS EST.

Norway rat.

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 479)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: meares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLVA=Yes.

FEATURES
 source location/Qualifiers
 1..479
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS2-bec-h-04-0-UI"
 /clone_11b="UI-R-BS2"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Bco RI; The UI-R-BS2 library is a subtracted library derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratel.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_L1b=UI-R-BS2
 TAG_TISSUE=embryo at 13 dpc
 TAG_SEQ=ATTC"
 BASE COUNT 145 a 92 c 108 g 134 t
 ORIGIN

Query Match 69.7%; Score 23; DB 12; Length 479;
 Best Local Similarity 83.9%; Pred. No. 6.9e+03;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TGTGTACTTCTCTCTATTCACGCCA 33
 383 TGTGTCTTTCTCTCTATTCACGACA 413

RESULT 5
 AK005882/c
 LOCUS
 DEFINITION AK005882 1522 bp mRNA linear HTC 19-JAN-2002
 Mus musculus adult male testis cDNA, RIKEN full-length cDNA library, clone:1700011N24; Retroviral aspartyl protease containing protein, full insert sequence.
 ACCESSION AK005882
 VERSION AK005882.1 GI:12838694
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain: C57BL/6J) adult male testis cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library
 clone:1700011N24.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, K., Ohara, E., Matsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, F., Gissi, C., King, B., Koculaw, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Wagner, L., Schriml, L.M., Staudt, J., Suzuki, R., Tomita, M., Madgler, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schobach, C., Seyer, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851
 REFERENCE 5 (bases 1 to 1522)
 AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hironaka, K., Hirooka, T., Hori, F., Hume, D., Inocenti, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kozawa, M., Koyama, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, T., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamuro, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Substitution
 JOURNAL Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp).

/clone="TA25042G03R"
 /clone_id="TA25"
 /cissue_type="developing seeds"
 /dev_stage="25 days after anthesis"
 /lab_host="E. coli DH10B"
 /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
 Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
 of cultivar Glenlea 25 days post-anthesis"
 BASE COUNT 270 a 111 c 148 g 168 t
 ORIGIN
 Query Match 68.5%; Score 22.6; DB 14; Length 697;
 Best Local Similarity 86.2%; Pred. No. 7.8e+03;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTTCTCTATTCAG 30
 Db 346 CTGTGTTCTTCTTCTCTCTCTAG 318
 RESULT 8
 BE876684/c 1527 bp mRNA linear EST 20-OCT-2000
 LOCUS 601488248P1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890914 5',
 DEFINITION mRNA sequence.
 ACCESSION BE876684
 VERSION BE876684.1 GI:10325460
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1527)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTF/Gardar
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: LMN674 row: 0 column: 11
 High quality sequence stop: 3.
 Location/Qualifiers
 1. 1527
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3890914"
 /clone_id="NIH_MGC_69"
 /cissue_type="large cell carcinoma, undifferentiated"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 BASE COUNT 800 a 393 c 324 g 10 t
 ORIGIN
 Query Match 68.5%; Score 22.6; DB 12; Length 1577;
 Best Local Similarity 86.2%; Pred. No. 6.2e+03;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTTCTCTATTCAG 30
 Db 1206 CTGTGTTGCTTCTTCTCTCTCTG 1178
 RESULT 9
 BH659639

LOCUS BH659639 693 bp DNA linear GSS 19-FEB-2002
 DEFINITION BOMEU91TF BO_2_3_KB Brassica oleracea genomic clone BOMEU91, DNA
 sequence.
 ACCESSION BH659639
 VERSION BH659639.1 GI:18718200
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 693)
 AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Contact: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.
 Location/Qualifiers
 1. 693
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOMEU91"
 /clone_id="BO_2_3_KB"
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHO51 using BstXI linkers"
 BASE COUNT 218 a 64 c 110 g 301 t
 ORIGIN
 Query Match 67.9%; Score 22.4; DB 17; Length 693;
 Best Local Similarity 81.2%; Pred. No. 8.8e+03;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTTCTCTATTCAGCA 33
 Db 146 CTGTGTTAGTTCTTCTCTATTCACAA 177
 RESULT 10
 AQ872730/c 455 bp DNA linear GSS 08-NOV-1999
 LOCUS AQ872730
 DEFINITION V15F1 mtn-3xHA/lacZ insertion library, strain AB972 Saccharomyces
 cerevisiae genomic 5', DNA sequence.
 ACCESSION AQ872730
 VERSION AQ872730.1 GI:6284876
 KEYWORDS GSS.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 455)
 AUTHORS Rose-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
 deBakges,S.A., Cheung,K.-H., Sheehan,A., Symonakis,D., Jensen,R.,
 Unanue,L., Heidman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
 Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: annj.kumar@yale.edu
 te of mtn-3xHA/lacZ insertion.

Seq primer: GGCTCTTCTTGTGAAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..455

/organism="Saccharomyces cerevisiae"
/strain="AB972 - trp1 r(0) (S288C background)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
/lab_host="E. coli"
/note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 159 a 104 c 72 g 120 t

Query Match 67.3%; Score 22.2; DB 17; Length 455;
Best Local Similarity 88.9%; Pred. No. 1.1e+04;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
|||||
Db 373 GCTGTGTGCTTCTTCTTGTGTC 347

RESULT 11

LOCUS AO874006 473 bp DNA linear GSS 08-NOV-1999
DEFINITION V969 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.

ACCESSION AO874006
VERSION AO874006.1 GI:6286250

KEYWORDS GSS.

SOURCE Baker's yeast.

ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 473)
Roes-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBerges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jensen, R., Unanueky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

Unpublished (1999)

JOURNAL

CONTACT: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu

te of mtn-3xHA/lacZ insertion.
Seq primer: GGCTCTTCTTGTGAAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..473

/organism="Saccharomyces cerevisiae"
/strain="AB972 - trp1 r(0) (S288C background)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
/lab_host="E. coli"
/note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 155 a 108 c 84 g 125 t 1 others

ORIGIN

Query Match 67.3%; Score 22.2; DB 17; Length 473;
Best Local Similarity 88.9%; Pred. No. 1.1e+04;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
|||||
Db 371 GCTGTGTGCTTCTTGTGTC 345

RESULT 12

LOCUS AO872889 577 bp DNA linear GSS 08-NOV-1999
DEFINITION V54C5 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.

ACCESSION AO872889
VERSION AO872889.1 GI:6285133

KEYWORDS GSS.

SOURCE Baker's yeast.

ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 577)
Roes-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBerges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jensen, R., Unanueky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

Unpublished (1999)

JOURNAL

CONTACT: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu

te of mtn-3xHA/lacZ insertion.
Seq primer: GGCTCTTCTTGTGAAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..577

/organism="Saccharomyces cerevisiae"
/strain="AB972 - trp1 r(0) (S288C background)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
/lab_host="E. coli"
/note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 180 a 147 c 105 g 145 t

ORIGIN

Query Match 67.3%; Score 22.2; DB 17; Length 577;
Best Local Similarity 88.9%; Pred. No. 1e+04;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
|||||
Db 325 GCTGTGTGCTTCTTGTGTC 299

RESULT 13

LOCUS AO872952 578 bp DNA linear GSS 08-NOV-1999
DEFINITION V56B3 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces cerevisiae genomic 5', DNA sequence.

ACCESSION AO872952
VERSION AO872952.1 GI:6285196

KEYWORDS GSS.

SOURCE Baker's yeast.

ORGANISM *Saccharomyces cerevisiae*
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 578)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
 deStranges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umanaky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumary@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCCTCTTCTTCTTGGAGAATAC
 Class: transposon-tagged.
 FEATURES
 source Location/Qualifiers
 1..578
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was
 prepared in pHS6-Sal; genomic DNA was size-fractionated
 (DNA of roughly 2-3 kb in length) prior to cloning. This
 library was subsequently mutagenized with a mtn-3xHA/lacZ
 mutitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 185 a 143 c 101 g 149 t
 ORIGIN
 Query Match 67.3%; Score 22.2; DB 17; Length 578;
 Best Local Similarity 88.9%; Pred. No. 1e+04;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GCTGTGTTACTTTCTTCTTCTATTC 27
 |||||
 Db 362 GCTGTGTTGCTTTCTTCTGTTCTC 336
 RESULT 14
 BG850976 708 bp mRNA linear EST 29-MAY-2001
 LOCUS 1024029H08.Y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BG850976
 VERSION BG850976.1 GI:14232160
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadales; Chlamydomonadae.
 1 (bases 1 to 708)
 AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
 McDermott, J.P., Silflow, C., Stern, D., and Surzycki, R.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hauser
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.
 Location/Qualifiers

source
 1..708
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
 II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in
 ambient levels of CO2 and HS medium bubbled with 5% CO2.
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into Lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the Lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."
 BASE COUNT 93 a 134 c 105 g 376 t
 ORIGIN
 Query Match 67.3%; Score 22.2; DB 12; Length 708;
 Best Local Similarity 88.9%; Pred. No. 9.8e+03;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTTCTTCTTCTATTC 28
 |||||
 Db 435 CTGTGTTATTTTCTTCTTCTATTC 461
 RESULT 15
 AO876136 810 bp DNA linear GSS 08-NOV-1999
 LOCUS V133f7 mtn-3xHA/lacZ Insertion library, strain Y2278 Saccharomyces
 cerevisiae genomic 5', DNA sequence.
 ACCESSION AO876136
 VERSION AO876136.1 GI:6288380
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 810)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
 deStranges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
 Umanaky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumary@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCCTCTTCTTCTTGGAGAATAC
 Class: transposon-tagged.
 FEATURES
 source Location/Qualifiers
 1..810
 /organism="Saccharomyces cerevisiae"
 /strain="Y2278 - S288C background, c1r(0) rho(0)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion library, strain Y2278"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 without 2 micron or mitochondrial DNA was prepared in
 pHS6-Sal; genomic DNA was size-fractionated (DNA of
 roughly 2-3 kb in length) prior to cloning. This library
 was subsequently mutagenized with a mtn-3xHA/lacZ

Mon May 12 14:15:08 2003

us-09-963-803-16.rst

Page 8

BASE COUNT	minitransposon containing lacZ, URA3, and tet resistance				
ORIGIN	"	a	c	g	t
	222	158	168	260	2 others

Query Match	67.3%	Score 22.2	DB 17	Length 810
Best Local Similarity	88.9%	Pred. No. 9.4e+03		
Matches 24	Conservative 0	Mismatches 3	Indels 0	Gaps 0

```

Qy      1  GCTGTGTACTTTCCTTCTATTC  27
          |||||  |||||  |||||  |||
Db     173  GCTGTGTGCTTTCCTTGTCTGTC  199

```

Search completed: May 11, 2003, 04:54:30
Job time : 535.461 secs

Page 8

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 60.2907 Seconds

(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-17

Sequence: 1 attccacccttaccattccgccttcattg 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	21	AAA96851
2	20.4	65.8	339	14	AA061194
3	20.4	65.8	1288	20	AAK61414
4	20.4	65.8	40962	23	ABL29926
5	20.4	65.8	2944528	24	ABA03041
6	19.8	63.9	498	24	ABL66989
7	19.8	63.9	3079	22	AAH4182
8	19.4	62.6	291	24	ABK76620
9	19.4	62.6	595	22	ABA61610

C 10	19.4	62.6	595	22	AA141521
C 11	19.4	62.6	1671	23	AA553289
C 12	19.4	62.6	1630121	17	AAT42063
C 13	19.2	61.9	537	22	AA914400
C 14	19.2	61.9	537	24	ABK37780
C 15	19.2	61.9	1000	22	AA913199
C 16	19.2	61.9	1000	24	ABK37779
C 17	19.2	61.9	1260	23	AA594538
C 18	19.2	61.9	1281	9	AA81257
C 19	19.2	61.9	1281	13	AA029507
C 20	19.2	61.9	1281	13	AA029508
C 21	19.2	61.9	1281	15	AA064903
C 22	19.2	61.9	1281	15	AA064904
C 23	19.2	61.9	1348	22	AA069554
C 24	19.2	61.9	2056	23	AA593259
C 25	19.2	61.9	2462	23	AA593184
C 26	19.2	61.9	2562	23	AA573488
C 27	19.2	61.9	4530	22	AA069558
C 28	19.2	61.9	30078	21	AA081520
C 29	19.2	61.9	49767	21	AA081458
C 30	19.2	61.9	56485	21	AA081476
C 31	19.2	61.9	172325	21	AA021613
C 32	19.2	61.9	349980	21	AA021608
C 33	19.2	61.9	349980	21	AA021612
C 34	19.2	61.9	837096	21	AA081489
C 35	19.2	61.9	1437668	21	AA081490
C 36	19.2	61.3	735	21	AA049868
C 37	19.2	61.3	811	21	AA049866
C 38	19.2	61.3	812	21	AA040364
C 39	19.2	61.3	1033	23	AA089565
C 40	19.2	61.3	1108	23	AA089561
C 41	19.2	61.3	1362	21	AA047934
C 42	19.2	61.3	14568	24	AB132230
C 43	19.2	61.3	23683	24	AB170482
C 44	19.2	61.3	23683	24	AB173423
C 45	18.8	60.6	319	22	ABA11402

ALIGNMENTS

RESULT 1	AAA96851	standard; DNA; 31 BP.
ID	AAA96851	
AC	AAA96851	
DT	19-FEB-2001	(first entry)
DE	Guide desoxynucleotide building block G3.	
XX	Promoter; intergenic region; Comelina yellow mottle virus;	
KM	chimeric expression promoter; plant vascular expression promoter;	
KM	plant green tissue expression promoter; Cassava vein mosaic virus;	
KM	transgenic plant; ss.	
XX	Synthetic.	
XX	WO200058485-A1.	
XX	05-OCT-2000.	
XX	29-MAR-2000; 2000WO-IB00370.	
XX	29-MAR-1999; 99FR-0003925.	
XX	(MERI-) MERISTEM THERAPEUTICS.	
XX	Rance I, Gruber V, Theisen M;	
XX	WPI; 2000-647238/62.	
PT	Chimeric expression promoter for transgenic plant production, comprises	

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 24; 91pp; English.
XX
CC The present sequence represents a guide desoxynucleotide building
CC block, which was used to link directional desoxynucleotide building
CC blocks during construction of chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comelia yellow mottle virus, and the second plant promoter originates
CC from the Casava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SO Sequence 31 BP; 5 A; 11 C; 2 G; 13 T; 0 other;
XX
Query Match 100.0%; Score 31; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATTCACCTTTTACATTCGCCGCTTCAATG 31
DB 1 ATTCACCTTTTACATTCGCCGCTTCAATG 31
XX
RESULT 2
ID AAG61194/c
AC AAG61194 standard; DNA; 339 BP.
XX
AC AAG61194;
XX
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST01232.
XX
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
PN W09316178-A.
XX
PD 19-AUG-1993.
XX
PF 12-FEB-1993; 93WO-US01234.
XX
PR 12-FEB-1992; 92US-0837195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Adams MD, Moreno RF, Venter CJ;
XX
DR WPI; 1993-272882/34.
XX
XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
PS Example 4; Page 452; 500pp; English.
XX
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.

CC EST01232 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AAG05041-Q61440.
XX
XX
SO Sequence 339 BP; 115 A; 66 C; 66 G; 92 T; 0 other;
XX
Query Match 65.8%; Score 20.4; DB 14; Length 339;
Best Local Similarity 80.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 2 TTTCACCTTTTACATTCGCCGCTTCAATG 31
DB 76 TTTCACCTTTTACATTCGCCGCTTCAATG 47
XX
RESULT 3
ID AAX61414
AC AAX61414 standard; cDNA; 1288 BP.
XX
AC AAX61414;
XX
DT 14-JUL-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS; ss.
XX
XX Homo sapiens.
OS
XX
XX W09922243-A1.
PN
XX
PD 06-MAY-1999.
XX
PF 23-OCT-1998; 98WO-US22376.
XX
XX 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Peng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kaye H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
XX WPI; 1999-303069/25.
DR P-PSDB; AAY19534.
XX
XX New isolated human genes and the secreted polypeptides they encode
PT
XX
PS Claim 3; Page 359; 546pp; English.
XX

CC The specification describes cDNA sequences (AAK61322-X61470) encoding
 CC human secreted proteins (AA19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukaemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.

XX Sequence 1288 BP; 333 A; 304 C; 237 G; 407 T; 7 other;

Query Match 65.8%; Score 20.4; DB 20; Length 1288;
 Best Local Similarity 80.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

2 TTCACCTTTACATCCCGCCTTCATTG 31
 1000 TACCACTTTAAATCCAGCTCATCATTTG 1029

RESULT 4

ABL29926/c
 ID ABL29926 standard; DNA; 40962 BP.

XX ABL29926;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 41251.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 41251; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB57072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 40962 BP; 12314 A; 8661 C; 8334 G; 11653 T; 0 other;

Query Match 65.8%; Score 20.4; DB 23; Length 40962;
 Best Local Similarity 95.5%; Pred. No. 70;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

7 CTTTACATTCGCGCTTCA 28

DB 11690 CTTTACATTCGCGCTTCA 11669

RESULT 5
 ABA03041/c
 ID ABA03041 standard; DNA; 2944528 BP.

XX ABA03041;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes EGD-e genome sequence.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

XX vitamin B12; bacterial infection; disease; ds.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

XX Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihl H, Dehoux P;

XX Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunet F, Cossart P;

XX Daniels J, Goebel W, Kneft J, Kuhn M, Ng E, Vazquez-Boland JA;

XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

XX Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;

XX Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -

XX Claim 1; SEQ ID NO 1; 192pp; French.

XX The present sequence is the genome sequence of Listeria monocytogenes
 CC EGD-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes and
 CC related organisms, and to study genetic polymorphisms and other genomes.
 CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms, and for biosynthesis and
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccine compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
 Query/Match 65.8%; Score 20.4; DB 24; Length 2944528;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 TTCGACCTTTACATTCGCGCTTCATTG 31
 DB 745445 TTCGACCTTTACATTCGCGCTTCATTG 745416

RESULT 6
 ABL66989/c
 ID ABL66989 standard; DNA; 498 BP.

AC ABL66989;
 XX
 DT 15-MAY-2002 (first entry)
 XX

DE Thyroid cancer related gene sequence SEQ ID NO:5326.
 XX
 KM Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KM cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
 KM gene; ds.

OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX

PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US10838.
 XX

XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-231333P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234099P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 22-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.

PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.

XX
 PA (AVALON PHARM.
 PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
 PI Soppe DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.

XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set
 XX
 PS Claim 1; SEQ ID 5326; 44pp; English.

XX
 CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.
 XX

SO Sequence 498 BP; 126 A; 85 C; 121 G; 164 T; 2 other;

Query Match 63.9%; Score 19.8; DB 24; Length 498;
 Best Local Similarity 77.4%; Pred. No. 60;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 ATTCGACCTTTACATTCGCGCTTCATTG 31
 DB 158 ATTCGACCTTTACATTCGCGCTTCATTG 128

RESULT 7
 AAH14182
 ID AAH14182 standard; cDNA; 3079 BP.

XX
 AC AAH14182;
 XX
 DT 26-JUN-2001 (first entry)
 XX

XX
 DE Human cDNA sequence SEQ ID NO:11429.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX

XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.

XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.

XX	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES INST.
PA	Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
P1	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	WPI, 2001-318749/34.
DR	
XX	primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PR	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
XX	
PS	Claim 8; SEQ ID 11429; 2537P + CD ROM; English.
XX	
CC	The present invention describes primer sets for synthesising 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAG95983 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
SQ	Sequence 3079 BP; 890 A; 664 C; 575 G; 950 T; 0 other;
XX	
Query Match	63.9%; Score 19.8; DB 22; Length 3079;
Best Local Similarity	77.4%; Pred. No. 81;
Matches 24; Conservative	0; Mismatches 7; Indels 0; Gaps 0
OY	1 ATTGCACCTTTACATTCGCCGCTTCATTG 31
DB	1217 ATACCATTTTTAAGTCCACCTTCAATG 1247
RESULT 8	
ABK76620	
ID	ABK76620 standard; DNA; 291 BP.
XX	
AC	ABK76620;
XX	
XX	
DT	13-AUG-2002 (first entry)
XX	
DE	Bacillus licheniformis genomic sequence tag (GST) #3911.
XX	
KW	Differential gene expression; genomic sequenced tag; GST;
KW	altered culture condition; environmental stress;
KW	physiological provocation; ds.
XX	
OS	Bacillus licheniformis.
XX	
PN	WO200229113-A2.
PD	
PD	11-APR-2002.
XX	
PF	05-OCT-2001; 2001WO-US31437.
XX	
XX	06-OCT-2000; 2000US-0680598.

PR	27-MAR-2001; 2001US-279526P.
XX	(NOVO) NOVOZYMES BIOTECH INC.
PA	(NOVO) NOVOZYMES AS.
XX	
PI	Berka R, Clausen IG;
XX	
DR	WPI; 2002-416684/44.
XX	
PT	Monitoring differential expression of several genes in first Bacillus
XX	cell relative to expression of same genes in one or more second
PT	Bacillus cells, by using substrate containing Bacillus genomic
PT	sequenced tag array -
PS	Claim 4; SEQ ID NO 3911; 200pp; English.
XX	
CC	The invention describes a method of monitoring differential expression of
CC	genes in a first Bacillus cell relative to expression of the genes in
CC	other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC	isolated from Bacillus cells to a substrate containing array of Bacillus
CC	genomic sequenced tags (GST), examining the array, and determining
CC	relative gene expression by an observed hybridisation reporter signal of
CC	a spot in the array. The method is useful for measuring the expression of
CC	genes in a first Bacillus cell relative to expression of the same genes
CC	in one or more second Bacillus cells. The method is useful for monitoring
CC	global expression of several genes from a Bacillus cell, discovering new
CC	genes, identifying possible functions of unknown open reading frames and
CC	monitoring gene copy number variation and stability. Monitoring changes
CC	in expression of genes may be used to provide a representation of the way
CC	in which Bacillus cells adapt to changes in culture conditions,
CC	environmental stress or other physiological provocation. Extensive
CC	follow-up characterisation is unnecessary, when one spot on an array
CC	equals one gene or one open reading frame, since sequence information is
CC	available. This sequence represents a genomic sequence tag (GST) used in
CC	the method of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at
CC	ftp.wipo.int/pub/published_pct_sequences.
SQ	
Sequence	291 BP; 96 A; 55 C; 39 G; 101 T; 0 other;
Query Match	62.6%; Score 19.4; DB 24; Length 291;
Best Local Similarity	79.3%; Pred. No. 81;
Matches	23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
CY	2 TTTCACCTTTACATTCGCGCCTTCAAT 30
DB	256 TTCCTCTTAATCATACCCTTCAT 284
RESULT 9	
ABA61610/c	
ID	ABA61610 standard; DNA; 595 BP.
AC	ABA61610;
XX	
DT	01-FEB-2002 (first entry)
XX	
DE	Human foetal liver single exon nucleic acid probe #9915.
XX	
KW	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157277-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00669.
XX	
PR	04-FEB-2000; 2000US-0180312.
XX	
PR	26-MAY-2000; 2000US-0207456.
XX	

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 9915; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 595 BP; 194 A; 89 C; 110 G; 202 T; 0 other;
Query Match 62.6%; Score 19.4; DB 22; Length 595;
Best Local Similarity 79.3%; Pred. No. 91;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCACCTTTTACATTCGCGCTTCATT 30
Db 64 TTCACCTTTCATCTCTGCTTTT 36
RESULT 10
AA141521/c
ID AA141521 standard; DNA; 595 BP.
XX
AC AA141521;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #10207 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 10207; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 595 BP; 194 A; 89 C; 110 G; 202 T; 0 other;
Query Match 62.6%; Score 19.4; DB 22; Length 595;
Best Local Similarity 79.3%; Pred. No. 91;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCACCTTTTACATTCGCGCTTCATT 30
Db 64 TTCACCTTTCATCTCTGCTTTT 36
RESULT 11
AAS53289/c
ID AAS53289 standard; DNA; 1671 BP.
XX
AC AAS53289;
XX
DT 13-FEB-2002 (first entry)
XX
DE Haemophilus influenzae DNA for cellular proliferation protein #71.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX Haemophilus influenzae.
XX
XX WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PF 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207272P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PDSB; AAU35430.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 6926; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programs. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1671 BP, 447 A, 276 C, 419 G, 529 T, 0 other;
Query Match 62.6%; Score 19.4; DB 23; Length 1671;
Best Local Similarity 79.3%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCACCTTTTACATTCGCCGCTTTCATT 30
DB 421 TTTCACCTTTTACATTCACACCTTGCTTT 393
RESULT 12
AAAT42063
ID AAAT42063 standard; DNA, 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KM Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UWJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
CC sequence at least 9% identical to (1). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.

SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
Query Match 62.6%; Score 19.4; DB 17; Length 1830121;
Best Local Similarity 79.3%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCACCTTTTACATTCGCCGCTTTCATT 30
DB 468917 TTTCACCTTTTACATTCACACCTTGCTTT 468945
RESULT 13
AAAF91400/C
ID AAFA91400 standard; DNA, 537 BP.
XX
AC AAFA91400;
XX
DT 04-MAY-2001 (first entry)
XX
DE N. meningitidis (serogroup B) PorB promoter region, SEQ ID:26.
XX
KM Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KM genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthel FJ, Dalemans WJ, Denoel P, Deguesne G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonard J, Voet P;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Disclosure; Page 83; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: Neisseria
CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC pneumoniae. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and

CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents a *Neisseria*
CC meningitidis promoter sequence.
XX
SQ Sequence 537 BP; 124 A; 102 C; 153 G; 158 T; 0 other;
Query Match 61.9%; Score 19.2; DB 22; Length 537;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
CY 1 ATTCACCTTTTACATCCCGCCT 24
DB 114 ATTCACGCTCTAGATCCCGCCT 91
RESULT 14
ABK37780/c
ID ABK37780 standard; DNA; 537 BP.
XX
AC ABK37780;
XX
DT 08-MAY-2002 (first entry)
XX
DE Strong promoter from PorB gene.
XX
KM Promot; de; Antibacterial; vaccine; bleb; Gram-negative bacteria;
KM outer membrane; LPS; lipopolysaccharide; meningitis; bacteraemia;
KM otitis media; pneumonia; chronic bronchitis; sinusitis.
XX
OS *Neisseria meningitidis* serogroup B.
XX
PN WO200209746-A2.
XX
PD 07-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-EP08857.
XX
PR 31-JUL-2000; 2000EP-0956369.
PR 08-FEB-2001; 2001GB-0003170.
XX
PA (SMIK) *SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans W, Denoel P, Dequeane G, Feron C, Garcon N;
PI Lobet Y, Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2002-188686/24.
XX
PT New immunogenic composition comprising an antigen derived from a
PT pathogen and a bleb preparation from *Neisseria meningitidis*, useful as
PT a vaccine for treating or preventing disease caused by the pathogen -
XX
PS Disclosure; Page 86; 125pp; English.
XX
CC The invention relates to an immunogenic composition comprising an antigen
CC derived from a pathogen capable of protecting a host against the
CC pathogen, mixed with an adjuvant comprising a bleb preparation derived
CC from a Gram-negative bacterial strain. The immunogenic composition
CC consists of *N. meningitidis* B blebs or *N. meningitidis* C polysaccharide
CC antigen. The blebs (derived from the outer membrane) may also
CC have their toxic lipopolysaccharide (LPS) content reduced using
CC heterologous down regulating sequences for LPS pathway genes or by
CC up regulating genes involved in LPS synthesis suppression, by
CC a promoter replacement technique. The immunogenic preparation is useful
CC in the manufacture of a medicament for the treatment of a disease caused
CC by the pathogen from which the antigen is derived (e.g. from
CC *Neisseria*, meningitis and bacteraemia, from *Moraxella*, otitis media and
CC pneumonia, and from *H. influenzae* chronic bronchitis, sinusitis,
CC pneumonia and otitis media). The bleb derived from *M. catarrhalis* or
CC from a non-typable *H. influenzae* is useful as an adjuvant in an
CC immunogenic composition comprising one or more pneumococcal capsular
CC polysaccharides or protein antigens. The present sequence is a strong
CC promoter suitable for constitutively expressing genes in

CC *N. meningitidis*.
XX
SQ Sequence 537 BP; 124 A; 102 C; 153 G; 158 T; 0 other;
Query Match 61.9%; Score 19.2; DB 24; Length 537;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
CY 1 ATTCACCTTTTACATCCCGCCT 24
DB 114 ATTCACGCTCTAGATCCCGCCT 91
RESULT 15
AAF91399/c
ID AAF91399 standard; DNA; 1000 BP.
XX
AC AAF91399;
XX
DT 04-MAY-2001 (first entry)
XX
DE *N. meningitidis* (serogroup A) PorB gene upstream sequence, SEQ ID:25.
XX
KM Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KM genetically modified; protective antigen expression; LPS detoxification;
KM LPS; lipid A; homologous recombination vector; immunisation;
KM immunoprotective; non-toxic; paediatric; cyclic; circular; ds.
XX
OS *Neisseria meningitidis*.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans WJ, Denoel P, Dequeane G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Claim 46; Page 83; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: *Neisseria*
CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Moraxella*
CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events

CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
 CC increase or decrease expression of that gene. Immunoprotective and
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
 CC are more immunogenic, less toxic and safer, and are particularly useful
 CC for paediatric use. The present sequence represents a specifically
 CC claimed *Neisseria meningitidis* nucleic acid sequence.

XX Sequence 1000 BP; 264 A; 223 C; 240 G; 273 T; 0 other;

Query Match 61.9%; Score 19.2; DB 22; Length 1000;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATTCACCTTTACATTCGCCCT 24
 |||||
 Db 114 ATTCACCTTTACATTCGCCCT 91

Search completed: May 11, 2003, 03:16:48
 Job time : 658.291 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 12.4311 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-17
Perfect score: 31
Sequence: 1 attccacatttaccatccgccttcattg 31

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents NA.*
2: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.8	63.9	856	US-08-998-416-308	Sequence 308, App
2	19.2	61.9	1348	US-09-064-693A-22	Sequence 22, Appl
3	19.2	61.9	4530	US-09-064-693A-26	Sequence 26, Appl
4	18.8	60.6	6693	US-08-961-527-195	Sequence 195, App
5	18.4	59.4	2465	US-08-619-812-5	Sequence 5, Appl
6	18.2	58.7	7070	US-08-619-554-3	Sequence 3, Appl
7	18	58.1	298	US-09-071-035-147	Sequence 147, App
8	18	58.1	441	US-09-071-035-145	Sequence 145, App
9	18	58.1	502	US-08-961-083-137	Sequence 137, App
10	18	58.1	25002	US-08-961-527-48	Sequence 48, Appl
11	17.8	57.4	685	US-09-328-111-101	Sequence 101, App
12	17.6	56.8	2765	US-09-327-885-2	Sequence 2, Appl
13	17.4	56.1	1202	US-09-058-489-29	Sequence 29, Appl
14	17.4	56.1	5300	US-08-938-669A-1	Sequence 1, Appl
15	17.4	56.1	5304	US-08-938-669A-2	Sequence 2, Appl
16	17.4	56.1	6169	US-08-938-669A-3	Sequence 3, Appl
17	17.2	55.5	658	US-09-470-191-57	Sequence 57, Appl
18	17.2	55.5	1485	US-08-976-259-13	Sequence 13, Appl
19	17.2	55.5	2337	US-09-134-001C-1031	Sequence 1031, Ap
20	17.2	55.5	3390	US-08-453-742-26	Sequence 26, Appl
21	17.2	55.5	3390	US-08-454-464-26	Sequence 26, Appl
22	17.2	55.5	3390	US-08-453-222-26	Sequence 26, Appl
23	17.2	55.5	3390	US-08-453-802-26	Sequence 26, Appl
24	17.2	55.5	3416	US-08-453-742-24	Sequence 24, Appl
25	17.2	55.5	3416	US-08-454-464-24	Sequence 24, Appl
26	17.2	55.5	3416	US-08-453-222-24	Sequence 24, Appl
27	17.2	55.5	3416	US-08-453-802-24	Sequence 24, Appl

C 28	17.2	55.5	3588	1	US-07-792-885A-2	Sequence 2, Appl
29	17.2	55.5	49795	4	US-09-453-702B-60	Sequence 60, Appl
30	17	54.8	522	2	US-08-909-965C-16	Sequence 16, Appl
31	17	54.8	755	4	US-09-020-956-21	Sequence 21, Appl
32	17	54.8	755	4	US-09-030-607-21	Sequence 21, Appl
33	17	54.8	755	4	US-09-605-785-21	Sequence 21, Appl
34	17	54.8	755	4	US-09-439-113-21	Sequence 21, Appl
35	17	54.8	755	4	US-09-352-616A-21	Sequence 21, Appl
36	17	54.8	755	4	US-09-322-149A-21	Sequence 21, Appl
37	17	54.8	2308	4	US-09-377-155-1	Sequence 1, Appl
38	17	54.8	2308	4	US-09-669-974-1	Sequence 1, Appl
39	17	54.8	3747	2	US-09-080-897-1	Sequence 1, Appl
40	17	54.8	3747	4	US-09-323-735-1	Sequence 1, Appl
41	17	54.8	5822	3	US-08-899-595-4	Sequence 4, Appl
42	17	54.8	5822	3	US-08-899-595-5	Sequence 5, Appl
43	17	54.8	45546	4	US-09-146-053-6	Sequence 6, Appl
44	16.8	54.2	1252	4	US-09-276-531-92	Sequence 92, Appl
45	16.8	54.2	1987	4	US-09-227-357-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-08-998-416-308
Sequence 308, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weig, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 308:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1251UP

US-08-998-416-308

Query Match 63.9%; Score 19.8; DB 4; Length 856;
Best Local Similarity 77.4%; Pred. No. 10;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATTCACCTTTTACATCCCGCTTCAATG 31
|||
Db 293 ATTACCCATTAAATTCGCGCTTTATG 323

RESULT 2

US-09-064-693A-22/c
Sequence 22, Application US/09064693A
Patent No. 6210937
GENERAL INFORMATION:
APPLICANT: Ward, Thomas E.
TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
ENGINEERED BACTERIA FOR PRODUCTION
OF A SPECIFIC PLASTICS PRECURSOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: W. Gary Goodson
ADDRESSEE: INEL--Lockheed Martin Idaho
ADDRESSEE: Technologies Co.
STREET: P.O. Box 1625
CITY: Idaho Falls
STATE: Idaho
COUNTRY: USA
ZIP: 83415-3810
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: Toshiba Satellite Pro T2150CDS
OPERATING SYSTEM: Windows95
SOFTWARE: Word Perfect 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,693A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: W. Gary Goodson
REGISTRATION NUMBER: 22,387
REFERENCE/DOCKET NUMBER: LIT-PI-296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (208)526-9469
TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1348 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-22

Query Match 61.9%; Score 19.2; DB 4; Length 1348;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 CTTTACATCCCGCTTCAATG 31
|||
Db 830 CTTTACAGTCCGCTTGAATG 807

RESULT 3

US-09-064-693A-26/c
Sequence 26, Application US/09064693A
Patent No. 6210937
GENERAL INFORMATION:
APPLICANT: Ward, Thomas E.
TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY

TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
OF A SPECIFIC PLASTICS PRECURSOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: W. Gary Goodson
ADDRESSEE: INEL--Lockheed Martin Idaho
ADDRESSEE: Technologies Co.
STREET: P.O. Box 1625
CITY: Idaho Falls
STATE: Idaho
COUNTRY: USA
ZIP: 83415-3810
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: Toshiba Satellite Pro T2150CDS
OPERATING SYSTEM: Windows95
SOFTWARE: Word Perfect 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,693A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: W. Gary Goodson
REGISTRATION NUMBER: 22,387
REFERENCE/DOCKET NUMBER: LIT-PI-296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (208)526-9469
TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 4530 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-26

Query Match 61.9%; Score 19.2; DB 4; Length 4530;
Best Local Similarity 87.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 CTTTACATCCCGCTTCAATG 31
|||
Db 2712 CTTTACAGTCCCGCTTGAATG 2689

RESULT 4

US-08-961-527-195
Sequence 195, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 6693 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-195

Query Match 60.6%; Score 18.8; DB 4; Length 6693;
Best Local Similarity 76.7%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTTACATTCGCCCTTTCATT 30
Db 492 ATTCCACCTTTTGATGACCCGCTTGAGTT 521

RESULT 5
US-08-619-812-5/C
Sequence 5, Application US/08619812
Patent No. 610066
GENERAL INFORMATION:
APPLICANT: POTTER, ANDREW A.
APPLICANT: THEISEN, MICHAEL
APPLICANT: HARLAND, RICHARD J.
APPLICANT: RIOUX, CLEMENT R.
TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: REED & ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,812
FILING DATE: 15-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/038,719
FILING DATE: 29-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0019.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 108..1850

US-08-619-812-5

Query Match 59.4%; Score 18.4; DB 3; Length 2465;
Best Local Similarity 78.6%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTTACATTCGCCCTTTCATT 28
Db 1220 ATTCCACATTTTACCTTGCCGACCTTTCA 1193

RESULT 6
US-08-619-554-3/C
Sequence 3, Application US/08619554
Patent No. 5821353
GENERAL INFORMATION:
APPLICANT: DOUGLAS, Cameron M.
APPLICANT: CHREBET, Gary L.
APPLICANT: CLEWAS, Joseph
APPLICANT: EL-SHERBINI, Mohammed
APPLICANT: FOOR, Forrest
APPLICANT: KAHN, Jennifer
APPLICANT: KELLY, Rosemarie, - PARENT, S.A.
APPLICANT: MARRINAN, Jean, - RAMADAN, N.M.
APPLICANT: MORIN, Nancy, - REGISTE, B.A.
APPLICANT: ONISHI, Janet, - SHEI, Gan-Ju
TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,554
FILING DATE: 01-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: COPPOLA, JOSEPH A.
REGISTRATION NUMBER: 38,413
REFERENCE/DOCKET NUMBER: 19104P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-6734
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 7070 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-619-554-3

Query Match 58.7%; Score 18.2; DB 1; Length 7070;
Best Local Similarity 74.2%; Pred. No. 70;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTTACATTCGCCCTTTCATTG 31
Db 46 ATTCCACCTTTTGACATGCGCATCTTTCAGTG 16

RESULT 7
US-09-071-035-147/C
Sequence 147: Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-147
Query Match 58.1%; Score 18; DB 4; Length 298;
Best Local Similarity 80.8%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 1 ATTCCACCTTTACATTCGCCCTTT 26
Db 249 ATTCCACCTTTAATTACCGCTTT 224
RESULT 8
US-09-071-035-145/C
Sequence 145: Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-145
Query Match 58.1%; Score 18; DB 4; Length 441;
Best Local Similarity 80.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 1 ATTCCACCTTTACATTCGCCCTTT 26
Db 392 ATTCCACCTTTAATTACCGCTTT 367
RESULT 9
US-08-961-083-137/C
Sequence 137: Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 502 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-137
Query Match 58.1%; Score 18; DB 3; Length 502;

Best Local Similarity 80.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 CCACCTTTACATCCGCGCTTTCAT 29
Db 420 CCACCTTGAGATGTCGCGCTTTCAT 395

RESULT 10

US-08-961-527-48/C
; Sequence 48; Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-48

Query Match 58.1%; Score 18; DB 4; Length 25002;
Best Local Similarity 80.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 CCACCTTTACATCCGCGCTTTCAT 29
Db 10612 CCACCTTGAGATGTCGCGCTTTCAT 10587

RESULT 11

US-09-328-111-101/C
; Sequence 101; Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Detti, Adnan
; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 685
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(685)
; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-101

Query Match 57.4%; Score 17.8; DB 4; Length 685;
Best Local Similarity 82.6%; Pred. No. 67;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 CCTTTACATCCGCGCTTTCAT 29
Db 527 CCNTGNCATTCGCCCTTTCAT 505

RESULT 12

US-09-397-885-2/C
; Sequence 2; Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schaefer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, -b-D-fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; EARLIER FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2769
; TYPE: DNA
; ORGANISM: Paenibacillus amylolyticus
; US-09-397-885-2

Query Match 56.8%; Score 17.6; DB 4; Length 2769;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 CACCTTTACATCCGCGCTTTCAT 28
Db 2584 CACATTCACATTCGCCCTTTTA 2561

RESULT 13

US-09-058-489-29

Sequence 29, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-089A
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 1202
TYPE: DNA
ORGANISM: Human
US-09-058-489-29

Query Match 56.1%; Score 17.4; DB 3; Length 1202;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATCCACCTTTACATTCGCCGCTTTC 27
DB 648 ACTCAACATTTACATTCATCTTTC 674

RESULT 14
US-08-938-669A-1/c
Sequence 1, Application US/08938669A
Patent No. 6171788
GENERAL INFORMATION:
APPLICANT: Nguyen, Thai D.
APPLICANT: Polansky, Jon R.
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS,
PROGNOSIS AND TREATMENT OF GLAUCOMA AND
TITLE OF INVENTION: RELATED DISEASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004-2402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,669A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/791,154
FILING DATE: 28-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mendelson, Elliot
REGISTRATION NUMBER: P-42,878
REFERENCE/DOCKET NUMBER: 07425-0034
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 383-6857
TELEFAX: 202 383-6610
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5300 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-938-669A-1

Query Match 56.1%; Score 17.4; DB 4; Length 5300;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCGCCGCTTCA 28
DB 2573 TTGCGCTTTACCTTCCTGCTTCA 2547

RESULT 15
US-08-938-669A-2/c
Sequence 2, Application US/08938669A
Patent No. 6171788
GENERAL INFORMATION:
APPLICANT: Nguyen, Thai D.
APPLICANT: Polansky, Jon R.
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS,
PROGNOSIS AND TREATMENT OF GLAUCOMA AND
TITLE OF INVENTION: RELATED DISEASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004-2402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,669A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/791,154
FILING DATE: 28-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mendelson, Elliot
REGISTRATION NUMBER: P-42,878
REFERENCE/DOCKET NUMBER: 07425-0034
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 383-6857
TELEFAX: 202 383-6610
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-938-669A-2

Query Match 56.1%; Score 17.4; DB 4; Length 5304;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCGCCGCTTCA 28
DB 2573 TTGCGCTTTACCTTCCTGCTTCA 2547

Search completed: May 11, 2003, 03:08:05
Job time : 16.4311 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 30.6892 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-17

Perfect score: 31

Sequence: 1 attccaccttaccattccgccttcattg 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues 1567708

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	9 US-09-963-803-17	Sequence 17, Appl
2	20.6	66.5	1691139	9 US-10-067-514-1	Sequence 1, Appl
3	19.8	63.9	387	10 US-09-960-352-14643	Sequence 14643, A
4	19.8	63.9	498	10 US-09-964-824A-23	Sequence 23, Appl
5	19.4	62.6	291	10 US-09-974-300-3911	Sequence 3911, Ap
6	19.4	62.6	595	10 US-09-864-761-15272	Sequence 15272, A
7	19.4	62.6	1671	10 US-09-815-242-6926	Sequence 6926, Ap
8	19.4	62.6	2000	9 US-09-938-842A-4351	Sequence 4351, Ap
9	19.4	62.6	501	10 US-09-783-590-526	Sequence 526, Ap
10	18.8	60.6	493	9 US-09-796-682-8653	Sequence 8653, Ap
11	18.8	60.6	493	9 US-10-040-862-8653	Sequence 8653, Ap
12	18.8	60.6	1038	10 US-09-822-849A-431	Sequence 431, Ap
13	18.8	60.6	32192	9 US-09-764-891-9791	Sequence 9791, Ap
14	18.8	60.6	32192	10 US-09-764-877-3657	Sequence 3657, Ap
15	18.6	60.0	32768	10 US-09-070-927A-399	Sequence 399, Ap
16	18.4	59.4	417	10 US-09-974-300-611	Sequence 611, Ap
17	18.4	59.4	468	9 US-10-079-854-96	Sequence 96, Appl
18	18.4	59.4	468	10 US-09-764-878-96	Sequence 96, Appl
19	18.4	59.4	580	10 US-09-560-863-525	Sequence 525, Appl

20	18.4	59.4	1039	7 US-08-781-986A-273	Sequence 273, App
21	18.4	59.4	1557	10 US-09-974-300-249	Sequence 249, App
22	18.2	58.7	108	10 US-09-974-300-7471	Sequence 7471, Ap
23	18.2	58.7	293	10 US-09-294-093B-3209	Sequence 3209, Ap
24	18.2	58.7	5230	10 US-09-838-718A-5	Sequence 5, Appl
25	18.2	58.7	5770	10 US-09-838-718A-7	Sequence 7, Appl
26	18.2	58.7	5870	10 US-09-838-718A-8	Sequence 8, Appl
27	18.2	58.7	5906	10 US-09-838-718A-6	Sequence 6, Appl
28	18.2	58.7	9412	9 US-10-032-393-51	Sequence 51, Appl
29	18.2	58.7	9412	9 US-10-032-393-52	Sequence 52, Appl
30	18.2	58.7	9417	9 US-10-032-393-48	Sequence 48, Appl
31	18.2	58.7	9425	9 US-10-032-393-46	Sequence 46, Appl
32	18.2	58.7	9437	9 US-10-032-393-17	Sequence 17, Appl
33	18.2	58.7	9484	9 US-10-032-393-49	Sequence 49, Appl
34	18.2	58.7	9566	9 US-10-032-393-50	Sequence 50, Appl
35	18.2	58.7	10929	9 US-10-032-393-7	Sequence 7, Appl
36	18.2	58.7	12733	9 US-10-032-393-47	Sequence 47, Appl
37	18.2	58.7	12739	9 US-10-032-393-8	Sequence 8, Appl
38	18.2	58.1	274	10 US-09-878-574-13572	Sequence 13572, A
39	18.2	58.1	502	10 US-09-765-272-137	Sequence 137, App
40	18.2	58.1	612	10 US-09-070-927A-832	Sequence 832, App
41	18.2	58.1	724	10 US-09-070-927A-931	Sequence 931, App
42	17.8	57.4	272	10 US-09-878-574-8758	Sequence 8758, Ap
43	17.8	57.4	362	9 US-09-796-692-6939	Sequence 6939, Ap
44	17.8	57.4	362	9 US-09-796-692-9539	Sequence 9539, Ap
45	17.8	57.4	362	9 US-10-040-862-6999	Sequence 6999, Ap

ALIGNMENTS

RESULT 1
US-09-963-803-17
Sequence 17, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 17
LENGTH: 31
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Guide desoxynucleotide building
US-09-963-803-17

Query Match 100.0%; Score 31; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 ATTCCACCTTTACATTCGCCCTTTCAATG 31
1 ATTCCACCTTTACATTCGCCCTTTCAATG 31

RESULT 2
US-10-067-514-1/c
Sequence 1, Application US/10067514
Publication No. US20030054531A1
GENERAL INFORMATION:
APPLICANT: Giletsdotcltr, Solveig
APPLICANT: Jonsdotcltr, Siff

APPLICANT: Revnisdottir, Sigridur Th.
TITLE OF INVENTION: HUMAN STROKE GENE
FILE REFERENCE: 2345.2010-003
CURRENT APPLICATION NUMBER: US/10/067,514
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: US 09/811/352
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1691139
TYPE: DNA
ORGANISM: Human
US-10-067-514-1

Query Match 66.5%; Score 20.6; DB 9; Length 1691139;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 CCACCTTTACATTCGCCGCTTCATT 30
Db 929591 CCACCTTTACATTCGCCGCTTCATT 929595

RESULT 3
US-09-960-352-14643
Sequence 14643, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathiasen, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 14643
LENGTH: 387
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 62-LIB34-082-Q1-E1-H10
US-09-960-352-14643

Query Match 63.9%; Score 19.8; DB 10; Length 387;
Best Local Similarity 77.4%; Pred. No. 66;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCCGCTTCATTG 31
Db 39 ATTCCACCTTTACATTCGCCGCTTCATTG 69

RESULT 4
US-09-964-824A-23/C
Sequence 23, Application US/09964824A
Patent No. US20020102531A1
GENERAL INFORMATION:
APPLICANT: Horigan, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-71
CURRENT APPLICATION NUMBER: US/09/964,824A
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US/60/236,033
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,032
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,028
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 583

SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 498
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(498)
OTHER INFORMATION: n=a,t,g or c
US-09-964-824A-23

Query Match 63.9%; Score 19.8; DB 10; Length 498;
Best Local Similarity 77.4%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCCGCTTCATTG 31
Db 158 ATTCCACCTTTACATTCGCCGCTTCATTG 128

RESULT 5
US-09-974-300-3911
Sequence 3911, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3911
LENGTH: 291
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-3911

Query Match 62.6%; Score 19.4; DB 10; Length 291;
Best Local Similarity 79.3%; Pred. No. 92;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TTCCACCTTTACATTCGCCGCTTCATT 30
Db 256 TTCCCTTTATCATACCCCTTCATT 284

RESULT 6
US-09-864-761-15272/C
Sequence 15272, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03


```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15272
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018469.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER. SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN PLACENTA. SIGNAL = 0.51
; US-09-864-761-15272

Query Match          62.6%; Score 19.4; DB 10; Length 595;
Best Local Similarity 79.3%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Cy 2 TTCACCTTTTACATTCGCCCTTTCATT 30
    |||||
Db 64 TTCCACCTTTTACATTCCTGCTTTT 36

RESULT 7
; US-09-815-242-6926/c
; Sequence 6926, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6926
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1671)
; US-09-815-242-6926

Query Match          62.6%; Score 19.4; DB 10; Length 1671;
Best Local Similarity 79.3%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Cy 2 TTCACCTTTTACATTCGCCCTTTCATT 30
    |||||
Db 421 TTTCACCTTTTACATTCACACCTTTCATT 393

RESULT 8
; US-09-938-842A-4351/c
; Sequence 4351, Application US/0938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4351
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-4351

Query Match          62.6%; Score 19.4; DB 9; Length 2000;
Best Local Similarity 79.3%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Cy 2 TTCACCTTTTACATTCGCCCTTTCATT 30
    |||||
Db 1704 TTTCACCTTTTACATTCCTTCCTTTCATT 1676

RESULT 9
; US-09-783-590-526/c
; Sequence 526, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
```

```

1 LOCATION: (416)
2 OTHER INFORMATION: n equals a,t,g, or c
3 NAME/KEY: misc feature
4 LOCATION: (420)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: misc feature
7 LOCATION: (421)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: misc feature
10 LOCATION: (422)
11 OTHER INFORMATION: n equals a,t,g, or c
12 NAME/KEY: misc feature
13 LOCATION: (430)
14 OTHER INFORMATION: n equals a,t,g, or c
15 NAME/KEY: misc feature
16 LOCATION: (433)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: misc feature
19 LOCATION: (438)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: misc feature
22 LOCATION: (444)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: misc feature
25 LOCATION: (452)
26 OTHER INFORMATION: n equals a,t,g, or c
27 NAME/KEY: misc feature
28 LOCATION: (455)
29 OTHER INFORMATION: n equals a,t,g, or c
30 NAME/KEY: misc feature
31 LOCATION: (460)
32 OTHER INFORMATION: n equals a,t,g, or c
33 NAME/KEY: misc feature
34 LOCATION: (463)
35 OTHER INFORMATION: n equals a,t,g, or c
36 NAME/KEY: misc feature
37 LOCATION: (476)
38 OTHER INFORMATION: n equals a,t,g, or c
39 NAME/KEY: misc feature
40 LOCATION: (480)
41 OTHER INFORMATION: n equals a,t,g, or c
42 NAME/KEY: misc feature
43 LOCATION: (484)
44 OTHER INFORMATION: n equals a,t,g, or c
45 US-09-783-590-526
46
47 Query Match 61.3%; Score 19; DB 10; Length 501;
48 Best Local Similarity 75.9%; Pred. No. 1.4e+02;
49 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
50
51 QY 2 TTCACCTTTTACATTCGCCCTTTCATT 30
52 ||||| ||||| ||||| |||||
53 Db 296 TTCACNTTTTAAACTGCANCCCTTGCAAT 268
54
55 RESULT 10
56 US-09-786-692-8653/C
57 Sequence 8653, Application US/09796692
58 Publication No. US20020198362A1
59 GENERAL INFORMATION:
60 APPLICANT: Gaiger, Alexander
61 APPLICANT: Algate, Paul A.
62 APPLICANT: Mannion, Jane
63 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
64 TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
65 FILE REFERENCE: 2077.001200
66 CURRENT APPLICATION NUMBER: US/09/796,692
67 PRIOR FILING DATE: 2001-03-01
68 PRIOR APPLICATION NUMBER: 60/186,126
69 PRIOR FILING DATE: 2000-03-01
70 PRIOR APPLICATION NUMBER: 60/190,479
71 PRIOR FILING DATE: 2000-03-17
72 PRIOR APPLICATION NUMBER: 60/200,545

```

? PRIOR FILING DATE: 2000-04-27
 ? PRIOR APPLICATION NUMBER: 60/200,303
 ? PRIOR FILING DATE: 2000-04-28
 ? PRIOR APPLICATION NUMBER: 60/200,779
 ? PRIOR FILING DATE: 2000-04-28
 ? PRIOR APPLICATION NUMBER: 60/200,999
 ? PRIOR FILING DATE: 2000-05-01
 ? PRIOR APPLICATION NUMBER: 60/202,084
 ? PRIOR FILING DATE: 2000-05-04
 ? PRIOR APPLICATION NUMBER: 60/206,201
 ? PRIOR FILING DATE: 2000-05-22
 ? PRIOR APPLICATION NUMBER: 60/218,950
 ? PRIOR FILING DATE: 2000-07-14
 ? PRIOR APPLICATION NUMBER: 60/222,903
 ? PRIOR FILING DATE: 2000-08-03
 ? PRIOR APPLICATION NUMBER: 60/223,416
 ? PRIOR FILING DATE: 2000-08-04
 ? PRIOR APPLICATION NUMBER: 60/223,378
 ? PRIOR FILING DATE: 2000-08-07
 ? NUMBER OF SEQ ID NOS: 9597
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO 8653
 ? LENGTH: 493

? TYPE: DNA
 ? ORGANISM: Homo sapiens
 ? FEATURE:
 ? NAME/KEY: unsure
 ? LOCATION: (117)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (158)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (165)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (225)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (234)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (252)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (279)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (308)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (332)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (386)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (389)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (465)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (486)
 ? OTHER INFORMATION: n=A,T,C or G
 ? US-09-796-692-8653

Query Match 60.6%; Score 18.8; DB 9; Length 493;
 Best local Similarity 74.2%; Pred. No. 1.7e+02;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 ATTCACCTTTTACATTCGCCCTTTTCATG 31
 | ||| ||||| | |||| ||||| ||

Db 226 ANACCATTTTAAAGTCCACCTTTTCATG 196
 RESULT 11
 US-10-040-862-8653/C
 ? Sequence 8653, Application US/10040862
 ? Publication No. US20030078396A1
 ? GENERAL INFORMATION:
 ? APPLICANT: Gaiger, Alexander
 ? APPLICANT: Gaiger, Paul A.
 ? APPLICANT: Mannion, Jane
 ? APPLICANT: Retter, Marc
 ? APPLICANT: Corixa Corporation
 ? TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
 ? FILE REFERENCE: 014058-013520US
 ? CURRENT APPLICATION NUMBER: US/10/040,862
 ? CURRENT FILING DATE: 2001-11-06
 ? PRIOR APPLICATION NUMBER: US 60/186,126
 ? PRIOR FILING DATE: 2000-03-01
 ? PRIOR APPLICATION NUMBER: US 60/190,479
 ? PRIOR FILING DATE: 2000-03-17
 ? PRIOR APPLICATION NUMBER: US 60/200,545
 ? PRIOR FILING DATE: 2000-04-27
 ? PRIOR APPLICATION NUMBER: US 60/200,303
 ? PRIOR FILING DATE: 2000-04-28
 ? PRIOR APPLICATION NUMBER: US 60/200,779
 ? PRIOR FILING DATE: 2000-04-28
 ? PRIOR APPLICATION NUMBER: US 60/200,999
 ? PRIOR FILING DATE: 2000-05-01
 ? PRIOR APPLICATION NUMBER: US 60/202,084
 ? PRIOR FILING DATE: 2000-05-04
 ? PRIOR APPLICATION NUMBER: US 60/206,201
 ? PRIOR FILING DATE: 2000-05-22
 ? PRIOR APPLICATION NUMBER: US 60/218,950
 ? PRIOR FILING DATE: 2000-07-14
 ? PRIOR APPLICATION NUMBER: US 60/222,903
 ? PRIOR FILING DATE: 2000-08-03
 ? PRIOR APPLICATION NUMBER: US 60/223,416
 ? PRIOR FILING DATE: 2000-08-04
 ? PRIOR APPLICATION NUMBER: US 60/223,378
 ? PRIOR FILING DATE: 2000-08-07
 ? PRIOR APPLICATION NUMBER: US 09/796,692
 ? PRIOR FILING DATE: 2001-03-01
 ? NUMBER OF SEQ ID NOS: 10467
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO 8653
 ? LENGTH: 493
 ? TYPE: DNA
 ? ORGANISM: Homo sapiens
 ? FEATURE:
 ? NAME/KEY: unsure
 ? LOCATION: (117)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (158)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (165)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (225)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (234)
 ? OTHER INFORMATION: n=A,T,C or G
 ? NAME/KEY: unsure
 ? LOCATION: (252)

```
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (279)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (332)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (386)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (389)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (465)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (486)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-8653
```

Query Match 60.6%; Score 18.8; DB 9; Length 493;

Best Local Similarity 74.2%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
QY 1 ATTCACCTTTTACATTCGCCCTTTCAATG 31
    ||||| ||||| ||||| ||||| |||||
Db 226 AMACCATTTTAAAGTCCACCTTCAATG 196
```

RESULT 12

```
US-09-822-849A-431/c
; Sequence 431, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fectel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Reenick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822-849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 431
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-431
```

```
Query Match 60.6%; Score 18.8; DB 10; Length 1038;
Best Local Similarity 76.7%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
QY 1 ATTCACCTTTTACATTCGCCCTTTCAAT 30
```

```
Db 519 AATCCCATTTTCCTTCCACCTTCATT 490
    ||||| ||||| ||||| ||||| |||||
```

RESULT 13

```
US-09-764-891-9791/c
; Sequence 9791, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9791
; LENGTH: 32192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9791
```

Query Match 60.6%; Score 18.8; DB 9; Length 32192;
Best Local Similarity 76.7%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
QY 1 ATTCACCTTTTACATTCGCCCTTTCAAT 30
    ||||| ||||| ||||| ||||| |||||
Db 955 ATTCTAATTTCTACTTCCAGCCTTACAT 926
```

RESULT 14

```
US-09-764-877-3657
; Sequence 3657, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3657
; LENGTH: 32192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3657
```

Query Match 60.6%; Score 18.8; DB 10; Length 32192;
Best Local Similarity 76.7%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
QY 1 ATTCACCTTTTACATTCGCCCTTTCAAT 30
    ||||| ||||| ||||| ||||| |||||
Db 31238 ATTCTAATTTCTACTTCCAGCCTTACAT 31267
```

RESULT 15

```
US-09-070-927A-399/c
; Sequence 399, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; APPLICANT: Patrick J. Dillon
```

```
US-09-070-927A-399/c
; Sequence 399, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; APPLICANT: Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
```

```

: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/070.927A
: FILING DATE: 04-May-2000
: CLASSIFICATION: <Unknown>
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/046,655
: FILING DATE: 1997-05-16
: APPLICATION NUMBER: 60/044,031
: FILING DATE: 1997-05-06
: APPLICATION NUMBER: 60/066,009
: FILING DATE: 1997-11-14
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Kenley K. Hoover
: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 399:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32768 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: US-09-070-927A-399

```

```

Query Match      60.0%; Score 18.6; DB 10; Length 32768;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATTCACCTTTTACATTCGCCGCTT 25
||||| ||||| ||||| |||||
Db 30220 ATTCACCTTTTACATTCGCCGCTT 30196

```

Search completed: May 11, 2003, 06:31:22
 Job time : 231.689 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 498.331 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-17

Perfect score: 31
Sequence: 1 attccacccttaccattcccgcttccatg 31

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_esthu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	71.0	548	13	BM318384 P11.13.A0
2	22	71.0	570	13	BM334924 BM334924
3	22	71.0	599	13	BM269265 BM269265
4	22	71.0	599	13	BM317907 BM317907
5	22	71.0	614	14	BM952936 BM952936
6	22	71.0	631	13	BM330506 BM330506

Result No.	Score	Query Match	Length	DB ID	Description
7	22	71.0	639	14	BQ283489 BQ283489
8	22	71.0	644	13	BM078660 BM078660
9	22	71.0	670	17	A2495987 A2495987
10	22	71.0	736	13	BM337587 BM337587
11	22	71.0	920	11	AY108329 AY108329
12	22	71.0	920	11	AY108329 AY108329
13	21.6	69.7	384	14	M79865 M79865
14	21.6	69.7	405	17	AQ500362 AQ500362
15	21.4	69.0	447	12	BG355209 BG355209
16	21.4	69.0	477	10	AW331076 AW331076
17	21.4	69.0	584	13	BI992430 BI992430
18	21.4	69.0	600	12	BG802971 BG802971
19	21.4	69.0	604	13	BM332582 BM332582
20	21.4	69.0	607	14	BQ547716 BQ547716
21	21.4	69.0	726	13	BM349877 BM349877
22	21.4	69.0	780	11	AY104156 AY104156
23	21.4	69.0	780	11	AY104156 AY104156
24	21.4	69.0	780	11	AY104156 AY104156
25	21.4	69.0	780	11	AY104156 AY104156
26	21.4	69.0	780	11	AY104156 AY104156
27	21.4	69.0	780	11	AY104156 AY104156
28	21.4	69.0	780	11	AY104156 AY104156
29	21.4	69.0	780	11	AY104156 AY104156
30	21.4	69.0	780	11	AY104156 AY104156
31	21.4	69.0	780	11	AY104156 AY104156
32	21.4	69.0	780	11	AY104156 AY104156
33	21.4	69.0	780	11	AY104156 AY104156
34	21.4	69.0	780	11	AY104156 AY104156
35	21.4	69.0	780	11	AY104156 AY104156
36	21.4	69.0	780	11	AY104156 AY104156
37	21.4	69.0	780	11	AY104156 AY104156
38	21.4	69.0	780	11	AY104156 AY104156
39	21.4	69.0	780	11	AY104156 AY104156
40	21.4	69.0	780	11	AY104156 AY104156
41	21.4	69.0	780	11	AY104156 AY104156
42	21.4	69.0	780	11	AY104156 AY104156
43	21.4	69.0	780	11	AY104156 AY104156
44	21.4	69.0	780	11	AY104156 AY104156
45	21.4	69.0	780	11	AY104156 AY104156

ALIGNMENTS

RESULT 1
LOCUS BM318384 548 bp mRNA linear EST 03-JAN-2002
DEFINITION P11.13.A02.b9.A002 Pathogen induced 1 (P11) Sorghum bicolor CDNA,
mRNA sequence.
ACCESSION BM318384
VERSION BM318384.1 GI:18052726
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 548)
Cordonnier-Pratt M.-W., Gingle A., Dean R., Sudhan M. and Pratt
L.H.
An EST database from Sorghum: pathogen-induced plants
unpublished (2000)
TITLE JOURNAL
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@arches.uga.edu

Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTmix or 7' sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV
High quality sequence stop: 447
POLYA-No.

FEATURES

Source Location/Qualifiers

1..548
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_1b="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pInescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Two-week-old sorghum plants (Brix 623 cultivar) were infected with pathogen (isolate FR421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 153 a 105 c 142 g 148 t

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 548;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATTG 31

Db 77 TTACACCTTACATCCCGCCTTAAATG 48

RESULT 2

BM334924

LOCUS BM334924 570 bp mRNA linear EST 16-JAN-2002

DEFINITION MEST130-A09.T3 ISUM5-RN Zea mays cDNA clone MEST130-A09 3', mRNA

ACCESSION BM334924 GI:18165085

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

TITLE 1 (bases 1 to 570)

JOURNAL Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.

COMMENT Including Sequence Tags from B73 Maize: various stages and tissues

Unpublished (2001)

Contact: Patrick S. Schnable

Schnable Laboratory

Iowa State University

G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Tel: 515-294-0975

Fax: 515-294-2299

Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the

FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES

Source Location/Qualifiers

1..570
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_1b="MEST130-A09"
/clone_1b="ISUM5-RN"
/tissue_type="mixed"
/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
/tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
(3, 5, 10, 15, 20, 25, 30, DAP). Adventitious roots (65 DAG
(3, 5, 10, 15, 20, 25, 30, DAP), Immature ear (0.2-3.0
(3, 5, 10, 15, 20, 25, 30, DAP), Immature ear (0.2-3.0
cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpolinated
first ear, ear shank, etiolated seedlings, callus,
Cycloheximide-treated callus, Anaerobic treated seedlings
, NAA (a-Naphthalene acetic acid)-treated seedlings,
kinetin-treated seedlings, ACP
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscisic acid
)-treated seedlings, GA (Gibberellic acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dT selected mRNA by priming with a
NotI oligo-dT primer (5'
AACTGGAAGAATTCGGCGCGCGCAATTTTCTTTTCTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA PolI-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcello Bento Soares (Genome
Research 6: 791-806, 1996)."

BASE COUNT 159 a 132 c 114 g 165 t

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 570;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATTG 31

Db 540 TTACACCTTACATCCCGCCTTAAATG 569

RESULT 3

BM269265

LOCUS BM269265 599 bp mRNA linear EST 18-DEC-2001

DEFINITION MEST407-A04.univ ISUM5-RN Zea mays cDNA clone MEST407-A04 3', mRNA

ACCESSION BM269265 GI:117932305

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

TITLE 1 (bases 1 to 599)

JOURNAL Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.

COMMENT Including Sequence Tags from B73 Maize: various stages and tissues

Unpublished (2001)

Contact: Patrick S. Schnable

Schnable Laboratory

Iowa State University

G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Tel: 515-294-0975
 Fax: 515-294-2299
 Email: schenab@iastate.edu
 Individual basical and confidence value were assigned using the
 pired software,
 (<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b>
 rt). Overall sequence quality assessment and vector trimming were
 conducted using the Lucy software (<http://www.tigr.org/sofftlab/>).
 Lucy parameters were set to ensure an overall trimmed quality of
 97.5% or better without any vector fragments in the chosen
 high-quality region of each sequence. Low-quality bases between the
 poly-T and the high-quality region were replaced with N's to serve
 as spacers.
 PCR Primers
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
 Seq primer: universal (GTA AAA CGA CGG CCA GT).
 Location/Qualifiers
 1. 599
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="MEST407-A04"
 /clone_1lb="ISUM5-RN"
 /issue_type="mixed"
 /lab_host="DH10B"
 /note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG
), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
 first ear, ear shank, etiolated seedlings, callus,
 Cycloheximide-treated callus, Anaerobic treated seedlings
 , NAA (a-Naphthalene acetic acid)-treated seedlings,
 Kinetin-treated seedlings, ACPG
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
 Brassinolide-treated seedlings, ABA (Abscissic acid
)-treated seedlings, GA (Gibberellic acid)-treated
 seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
 molecules were generated as follows. First-strand cDNA was
 prepared from oligo-dT selected mRNA by priming with a
 NotI oligo-dT primer (5',
 ACTGGAAGATTTCGGCCGCGAGATTTTCTTTT). The
 resulting DNA:RNA hybrid was treated with RNase H and used
 as a template for DNA Pol-I-catalyzed second strand
 synthesis. After the addition of EcoRI adaptor, the
 ds-cDNAs were digested with NotI and size-selected. The
 resulting molecules were directionally cloned into the
 EcoRI and NotI sites of the pT73PAC vector. The library
 then went through one round of normalization to COT value
 of 5 based on the methods of Marcelo Bento Soares (Genome
 Research 6: 791-806, 1996)."
 Research 6: 791-806, 1996)."
 147 c 115 g 178 t

BASE COUNT 159 a 147 c 115 g 178 t
 ORIGIN
 Query Match 71.0%; Score 22; DB 13; Length 599;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 2 TTCACCTTTTACATTCGCGCTTCATTG 31
 |||||
 Db 518 TTACACCTTAAACATCCCGCTTAATTG 547
 RESULT 4 599 bp mRNA linear EST 03-JAN-2002
 BM317907
 LOCUS
 DEFINITION P11.13_A02.g9_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BM317907
 VERSION BM317907.1 GI:18052249
 KEYWORDS EST.

SOURCE
 ORGANISM
 sorghum.
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 599)
 Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: pathogen-induced plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude polyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20. Three-prime sequences, which are obtained with polyTmix or
 T7 sequencing primer, are presented as the reverse complement.
 Seq primer: T7
 High quality sequence start: 60
 High quality sequence stop: 599
 POLYA=Yes.
 Location/Qualifiers
 1. 599
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_1lb="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from Lambda Zap II; Site 1: XhoI;
 Site 2: EcoRI; Two-week-old sorghum plants (BRX 623
 cultivar) were infected with pathogen (isolate FRW421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: Young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."
 117 c 142 g 170 t
 BASE COUNT 170 a 117 c 142 g 170 t
 ORIGIN
 Query Match 71.0%; Score 22; DB 13; Length 599;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 2 TTCACCTTTTACATTCGCGCTTCATTG 31
 |||||
 Db 44 TTACACCTTAAACATCCCGCTTAATTG 15
 RESULT 5 614 bp mRNA linear EST 14-MAR-2002
 BM952936
 LOCUS
 DEFINITION 952060C11.X1.952 - BMS tissue from Walbot lab (reduced RNA) Zea
 mays cDNA, mRNA sequence.
 ACCESSION BM952936
 VERSION BM952936.1 GI:19436526
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 614)

AUTHORS Walbot, V.
TITLE maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 952060 row: C column: 11.

FEATURES

Source

1. .614
 /organism="Zea mays"
 /cultivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:45577"
 /clone_lib="952 - BMS tissue from Walbot Lab (reduced RNA)"
 /tissue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth phases"
 /lab_host="DH10B"
 /note="Vector: pUC19, Site_1: EcoRI, Site_2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT 175 a 144 c 150 g 145 t

ORIGIN

Query Match 71.0%; Score 22; DB 14; Length 614;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTCACCTTTTACATCCCGCTTTCATTG 31

Db 567 TTACACCTTACATCCCGCTTTCATTG 538

RESULT 6 BM330506/c 631 bp mRNA linear EST 04-JAN-2002
LOCUS PIC1_52_H06.g1_A002 Pathogen-infected compatible 1 (PIC1) Sorghum
DEFINITION bicolor cDNA, mRNA sequence.
ACCESSION BM330506
VERSION BM330506.1 GI:18069643
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 631)
 Cordonnier-Pratt, M.-M., Gingle, A., Pang, G.C., Dean, R., Wing, R.,
 Sudman, M. and Pratt, L.H.
 An EST database from Sorghum: plants infected with a compatible
 pathogen
 Unpublished (2002)
JOURNAL Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmprratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector, and regions
 below Phred quality 16. The threshold for highest quality sequence

is 20. Three-prime sequences, which are obtained with PolyTmX or
 T7 sequencing primer, are presented as the reverse complement.
 Seq primer: T7
 High quality sequence start: 100
 High quality sequence stop: 631
 POLYA=yes.

FEATURES

Source

1. .631
 /organism="Sorghum bicolor"
 /cultivar="Brx623"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen-infected compatible 1 (PIC1)"
 /tissue_type="Leaves"
 /dev_stage="4-week-old seedlings infected with
 Colletotrichum graminicola"
 /note="Vector: pBluescript II SK(-) from Lambda Zap II;
 Site_1: XhoI, Site_2: EcoRI; Four-week-old sorghum
 seedlings were sprayed with spore suspension prepared from
 3-week-old PM421, a sorghum isolate of the anthracnose
 pathogen Colletotrichum graminicola. Inoculated plants
 were kept in a 25 C dark growth chamber with 100% relative
 humidity for 24 hr, followed by 12/12 hr of light/dark
 cycle at 25 C with 90% relative humidity for another 24
 hr. All leaves were harvested and quick frozen with liquid
 nitrogen and stored in a -80 C freezer. The library was
 made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision.
 WARNING: While most or all ESTs are expected to derive
 from the host plant, no effort was made to eliminate ESTs
 deriving from the pathogen."

BASE COUNT 181 a 120 c 154 g 176 t

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 631;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTCACCTTTTACATCCCGCTTTCATTG 31

Db 94 TTACACCTTACATCCCGCTTTCATTG 65

RESULT 7 BQ283489/c 639 bp mRNA linear EST 13-MAY-2002
LOCUS WHE3091_G03_M05ZS Wheat cold-stressed seedling subtracted cDNA
DEFINITION library Triticum aestivum cDNA clone WHE3091_G03_M05, mRNA
 sequence.
ACCESSION BQ283489
VERSION BQ283489.1 GI:20553240
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Triticum.
 1 (bases 1 to 639)
 Anderson, O.D., Cho, S., Choi, D.W., Close, T.J., Crossman, C., Fenton
 , R.D., Iazo, G.R., Nguyen, H.T., Phan, J., Rauech, C.J., Wilson, C., Woo
 , J. and Zhang, D.
 The structure and function of the expressed portion of the wheat
 genomes - Cold-stressed seedling subtracted cDNA library
 Unpublished (2002)
JOURNAL Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105959773
 Fax: 5105959818
 Email: oanderson@pw.usda.gov
 Sequences have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: SK primer.

FEATURES

Source

Location/Qualifiers
 1. .639
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE3091_G03_M05"
 /clone_1lb="Wheat cold-stressed seedling subtracted cDNA library"

/tissue type="Seedling"
 /dev stage="Five-day old seedling"
 /lab_host="E. coli 80Lr"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, mycstatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give Bluescript phagemids before subtraction was carried out. The mass excision of phagemid library and subtraction were done in HT Nguyen lab by D. Zhang at Texas Tech University. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)"

BASE COUNT 192 a 132 c 157 g 157 t 1 others
 ORIGIN

Query Match 71.0%; Score 22; DB 14; Length 639;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCGCTTCAATG: 31

Db 301 TTACACCTTACATCCCGCTTAAATG 272

RESULT 8

LOCUS

BM078660 644 bp mRNA linear EST 14-NOV-2001

DEFINITION MEST122-F07.T3 ISUM4-TN Zea mays cDNA clone MEST122-F07 3', mRNA

ACCESSION BM078660

VERSION BM078660.1 GI:16925592

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE

AUTHORS Qiu, F., Cui, F., Guo, L., Ashlock, D.A., Wen, T.J. and Schnable, P.S.

TITLE Expressed Sequence Tags from B73 Maize Seedlings and Silks

JOURNAL Unpublished (2001)

COMMENT Contact: Patrick S. Schnable

Schnable Laboratory
 Iowa State University
 G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
 Tel: 515-294-0975
 Fax: 515-294-2299
 Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the Phred software,
 (<http://depts.washington.edu/ventures/collabr/direct/index.htm#b>
 rt). Overall sequence quality assessment and vector trimming were
 conducted using the Lucy software (<http://www.tigr.org/softlab/>).
 Lucy parameters were set to ensure an overall trimmed quality of
 97.5% or better without any vector fragments in the chosen
 high-quality region of each sequence. Low-quality bases between the
 poly-T and the high-quality region were replaced with N's to serve

as spacers.

PCR Primers

FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA

CMA TAG)

BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC

TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES

Source

Location/Qualifiers
 1. .644

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="MEST122-F07"

/clone_1lb="ISUM4-TN"

/tissue type="Seedling and silk"

/lab_host="DH10B"

/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
 ds-cDNA molecules were generated as follows. First-strand
 cDNA was prepared from oligo-dT selected mRNA by priming
 with a NotI oligo-dT primer (5',
 AACTGAGAGATTCGGCGCGGCGGAGATTTTTTTTTTTTTTTT). The
 resulting DNA:RNA hybrid was treated with RNase H and used
 as a template for DNA Pol-I-catalyzed second strand
 synthesis. After the addition of EcoRI adaptors, the
 ds-cDNAs were digested with NotI and size-selected. The
 resulting molecules were directionally cloned into the
 EcoRI and NotI sites of the pT73PAC vector. The library
 then went through one round of normalization to COT value
 of 5 based on the methods of Marcelo Bento Soares (Genome
 Research 6: 791-806, 1996)."

BASE COUNT 168 a 154 c 121 g 201 t
 ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 644;
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCGCTTCAATG 31

Db 485 TTACACCTTACATCCCGCTTAAATG 514

RESULT 9

LOCUS

AZ495987 670 bp DNA linear GSS 05-OCT-2000

DEFINITION IM0332J11F Mouse 10kb plasmid UGCGIM library Mus musculus genomic

clone UGCGIM0332J11 F, DNA sequence.

ACCESSION AZ495987

VERSION AZ495987.1 GI:10671822

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: rdunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0332 row: J column: 11

Seq primer: CGTTGTAAACGACGCGCAGT

Class: plasmid ends
High quality sequence stop: 670.
Location/Qualifiers

FEATURES

Source

1. 670
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCM032311"
/clone_lib="Mouse 10kb plasmid UGCM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42mv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

217 a 124 c 121 g 208 t

ORIGIN

Query Match 71.0%; Score 22; DB 17; Length 670;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 TTCACCTTTACATTCGCCCTTCATTG 31
|||||
Db 536 TTCACCTTTACTTCTGCTTCCACTG 565

RESULT 10

BM337587
LOCUS

DEFINITION BM337587 736 bp mRNA linear EST 16-JAN-2002
MEST209-B11.T3 ISUM5-RN Zea mays cDNA clone MEST209-B11 3', mRNA
sequence.

ACCESSION BM337587
VERSION BM337587.1 GI:18167747

KEYWORDS EST.

SOURCE

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 736)

Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones

Unpublished (2001)
Contact: Patrick S. Schnable

JOURNAL

COMMENT

Schnable Laboratory
Iowa State University
GA05 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
Tel: 515-294-0975
Fax: 515-294-2299

Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
phred software.

(<http://depts.washington.edu/ventures/collabr/direct/index.htm#b>
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlib/lucy/>).
Lucy parameters were set to ensure an overall trimmed quality of

97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve
as spacers.

PER PRIMERS

FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)

BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

Location/Qualifiers

FEATURES

Source

1. 736
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST209-B11"
/clone_lib="ISUM5-RN"
/tissue_type="mixed"
/lab_host="DH10B"
/note="Vector: pT73PAC; Site 1: EcoRI, Site 2: NotI.
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG
, Tassel (3-39 cm, 53 and 56 DAG), Silk, unpolinated
cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpolinated
first ear, ear shank, etiolated seedlings, callus,
Cycloheximide-treated callus, Anaerobic treated seedlings,
NAA (a-Naphthalene acetic acid)-treated seedlings,
Kineth-treated seedlings, ACP
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscissic acid
) -treated seedlings, GA (Gibberellin acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dT selected mRNA by priming with a
NotI oligo-dT primer (5'
ACTGGAAGATTCCGCCCGCAGCAATTTTCTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to COT value
of 5 based on the methods of Marcello Bento Soares (Genome
Research 6: 791-806, 1996)."

BASE COUNT 181 a 164 c 146 g 225 t 20 others

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 736;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 TTCACCTTTACATTCGCCCTTCATTG 31
|||||
Db 538 TTCACCTTTACATTCGCCCTTCATTG 567

2 TTCACCTTTACATTCGCCCTTCATTG 31

RESULT 11

AY108329/c

LOCUS AY108329 920 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays PCO128508 mRNA sequence.
ACCESSION AY108329
VERSION AY108329.1 GI:21211407

KEYWORDS

HTC.

SOURCE

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 920)

Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitelc, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.

Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes

JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 920)
 AUTHORS Coe,B.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 FEATURES Location/Qualifiers
 SOURCE 1..920
 /organism="Zea mays"
 /db_xref="taxon:4577"
 /db_xref="maizedb:637475"
 /clone="PCO128508"
 /clone_lib="Maize Mapping Project/DuPont Cornsensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 230 a 247 c 224 g 213 t 6 others
 ORIGIN

Query Match 71.0%; Score 22; DB 11; Length 920;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATG 31
 Db 763 TTACACCTTTAACATCCCGCTTAAATG 734

RESULT 12
 LOCUS M79865 384 bp mRNA linear EST 30-JUN-1992
 DEFINITION WIST00402 Mixed stage, StrataGene (cat. #937006) Caenorhabditis
 elegans CDNA clone CEMSE75, mRNA sequence.
 M79865
 M79865.1 GI:271884
 EST.
 Caenorhabditis elegans.
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 384)
 McCombie,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G., Uterback
 T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and Fields,C.
 Caenorhabditis elegans expressed sequence tags identify gene
 families and potential disease gene homologues
 Nat. Genet. 1, 124-131 (1992)
 JOURNAL 93250983
 MEDLINE
 COMMENT Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..384
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="CEMSE75"
 /clone_lib="Mixed stage, StrataGene (cat. #937006)"
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage CDNA
 library. StrataGene catalog #937006. The library is oligo
 dt primed and directionally cloned in the Uni-ZAP XR
 vector."

BASE COUNT 156 a 46 c 74 g 101 t 7 others
 ORIGIN

Query Match 69.7%; Score 21.6; DB 14; Length 384;

Best Local Similarity 80.0%; Pred. No. 2.1e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 ATTCCACCTTTTACATTCGCCCTTTCAT 30
 Db 338 ATTCCACCTTTTAAATTCGCCCTTTAAT 309

RESULT 13
 LOCUS A0500362 605 bp DNA linear GSS 29-APR-1999
 DEFINITION Y41F1 5'HA/lacZ insertion library Saccharomyces cerevisiae
 genomic 5'HA, DNA sequence.
 A0500362
 A0500362.1 GI:4706072
 GSS.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 605)
 Rose-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
 deBerges,S.A., Cheung,K.-H., Sheehan,A., Symonakis,D., Hansen,R.,
 Umaneky,L., Heidman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
 Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 CONTACT: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumare@yale.edu
 Seq primer: GGCCTTCCTTCCTTCGACAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 1..605
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ insertion library"
 /lab_host="S. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 (lacking mitochondrial DNA) was prepared in pHS6-Sal;
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb
 in length) prior to cloning. This library was
 subsequently mutagenized with a mtn-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."

FEATURES
 SOURCE

BASE COUNT 160 a 149 c 108 g 188 t
 ORIGIN

Query Match 69.7%; Score 21.6; DB 17; Length 605;
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCAT 29
 Db 578 TTCACCTTTTAAATTCGCCCTTTAAT 605

RESULT 14
 LOCUS BG355209 447 bp mRNA linear EST 06-MAR-2001
 DEFINITION 947040C05.y2 947 - 2 week shoot from Barikan lab Zea mays CDNA, mRNA
 sequence.
 BG355209
 BG355209.1 GI:13237195
 EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
AUTHORS	clade; Panicoidae; Andropogoneae; Zea.
TITLE	1 (bases 1 to 447)
JOURNAL	Walbot, V.
COMMENT	Maize ESTs from various cDNA libraries sequenced at Stanford University Unpublished (1999) Contact: Walbot V

BASE COUNT	137 a	78 c	112 g	120 f
ORIGIN				

	Query Match	69.0%	Score 21.4	DB 12	Length 447
	Best Local Similarity	80.6%	Pred. No. 2.6e+02		
	Matches 25	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Oy	1 ATTCACCTTTTACATGCCGCTTGATTG	31			
Dh	334 ATTCTCCTTTACTTCTTGCCCTTCCCTG	304			

[illegible]

```

FEATURES
source
Location/Qualifiers
1. .477
/organism="Zea mays"
/cultivar="W23"
/db xref="taxon:4577"

```

```

/clone_lib=707 Mixed adult tissues from Waldbot lab (SK
)
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: tassel, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site 1: ECORI; cDNA library from fully
differentiated maize tissues from an active Mutator
plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."
BASE COUNT      155 a      70 c      116 g      136 t
ORIGIN

```

BASE COUNT	155 a	70 c	116 g	136 t	
ORIGIN					
Query Match		69.0%	Score 21.4;	DB 10;	Length 477;
Best Local Similarity		80.6%	Pred. No. 2.7e+02;		
Matches 25;	Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0

```

Query Match      69.0% ; Score 21.4 ; DB 10 ; Length 477 ;
Best Local Similarity 80.6% ; Pred. No. 2.7e+02 ;
Matches 25 ; Conservative 0 ; Mismatches 6 ; Indels 0 ; Gaps 0 ;

QY      1 ATTGCACCTTTTACATTCGCCGCTTTCAATG 31
          ||| ||| ||| ||| ||| ||| ||| |||
DB      163 ATTTCCTCTTACTCTTCTGCTCTTCTCTG 133

```

Search completed: May 11, 2003, 04:54:34
Job time : 502.331 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 338.927 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32

Sequence: 1 caagggtcgaataaagaataattcgtga 32

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapekt 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280.

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : *

1: gb Da:*

2: gb Htg:*

3: gb In:*

4: gb Om:*

5: gb Ov:*

6: gb Pat:*

7: gb Ph:*

8: gb Pl:*

9: gb Pr:*

10: gb Ro:*

11: gb Sts:*

12: gb Sy:*

13: gb Un:*

14: gb Vi:*

15: em Ba:*

16: em Fun:*

17: em Hum:*

18: em In:*

19: em Mu:*

20: em Om:*

21: em Or:*

22: em Ov:*

23: em Pat:*

24: em Ph:*

25: em Pl:*

26: em Ro:*

27: em Sts:*

28: em Un:*

29: em Vi:*

30: em Htg_Hum:*

31: em Htg_Inv:*

32: em Htg_Other:*

33: em Htg_Mus:*

34: em Htg_Pln:*

35: em Htg_Rod:*

36: em Htg_Mam:*

37: em Htg_Vit:*

38: em Sy:*

39: em Htgo_Hum:*

40: em Htgo_Mus:*

41: em Htgo_Other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	100.0	32	6	AX036752	AX036752 Sequence
2	22.4	70.0	116498	2	OSJ000004	AL606449 Oryza sat
3	22.4	70.0	167191	2	AC123085	AC123085 Rattus no
4	22.4	70.0	208214	10	AL645908	AL645908 Mouse DNA
5	22	68.8	213383	10	AL645943	AL645943 Mouse DNA
6	21.4	66.9	41502	3	CEZK856	Z70783 Caenorhabdi
7	21.4	66.9	158886	2	AC016944	AL161899 Human DNA
8	21.4	66.9	159981	2	AC025725	AC025725 Caenorhab
9	21.4	66.9	188215	2	AC084158	AC084158 Caenorhab
10	21.4	66.9	188616	3	AC093418	AC093418 Homo sapi
11	21.4	66.9	206320	9	AC006892	AC006892 Caenorhab
12	21.4	66.9	299081	2	AR016866	AR016866 Sequence
13	21.2	66.2	1567	6	AR020892	AR020892 Sequence
14	21.2	66.2	1567	6	AR027215	AR027215 Sequence
15	21.2	66.2	1567	6	AR027215	AR027215 Sequence
16	21.2	66.2	1567	6	AR038502	AR038502 Sequence
17	21.2	66.2	1567	6	AR064644	AR064644 Sequence
18	21.2	66.2	1567	6	AR067569	AR067569 Sequence
19	21.2	66.2	1567	6	I38521	I38521 Sequence 10
20	21.2	66.2	1567	6	I56996	I56996 Sequence 10
21	21.2	66.2	1567	6	I59862	I59862 Sequence 10
22	21.2	66.2	1567	6	I75189	I75189 Sequence 10
23	21.2	66.2	94727	9	AL591034	AL591034 Human DNA
24	21.2	66.2	114468	9	AC093026	AC093026 Homo sapi
25	21.2	66.2	162808	2	AC021340	AC021340 Homo sapi
26	21.2	66.2	145528	2	AC016800	AC016800 Homo sapi
27	21.2	66.2	169161	2	AC019053	AC019053 Homo sapi
28	21.2	66.2	171490	2	AC068209	AC068209 Homo sapi
29	21	65.6	2898	6	AX010976	AX010976 Sequence
30	21	65.6	2898	6	PFU278385	PFU278385 Penicilli
31	21	65.6	229061	9	AC007786	AC007786 Homo sapi
32	20.8	65.0	3103	1	AF259420	AF259420 Staphyloc
33	20.8	65.0	3103	6	AX144740	AX144740 Sequence
34	20.8	65.0	3438	1	AF270201	AF270201 Staphyloc
35	20.8	65.0	3438	6	AX145519	AX145519 Sequence
36	20.8	65.0	3463	1	AF269957	AF269957 Staphyloc
37	20.8	65.0	3463	6	AX145275	AX145275 Sequence
38	20.8	65.0	38621	2	AC079778	AC079778 Homo sapi
39	20.8	65.0	55835	2	AL138772	AL138772 Human DNA
40	20.8	65.0	100886	2	AP003831	AP003831 Oryza sat
41	20.8	65.0	111500	9	AC121344	AC121344 Homo sapi
42	20.8	65.0	113071	2	CNS07YPK	AL713938 Oryza sat
43	20.8	65.0	115393	8	AC091247	AC091247 Oryza sat
44	20.8	65.0	117505	8	AC096687	AC096687 Oryza sat
45	20.8	65.0	122381	9	AP000577	AP000577 Homo sapi

ALIGNMENTS

RESULT 1

AX036752

LOCUS

DEFINITION Sequence 18 from Patent WO0058485.

ACCESSION AX036752

VERSION AX036752.1 GI:11226261

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 32)

AUTHORS

Rance, I., Theisen, M. and Gruber, V.

Chimeric expression promoters originating from commelina yellow

motile virus and cassava vein mosaic virus

PATENT: WO 0058485-A 18 05-OCT-2000;

JOURNAL

linear PAT 16-NOV-2000

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
 ; GRUBER VERONIQUE (FR)
 FEATURES
 LOCATION/Qualifiers
 1. 32
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Guide desoxynucleotide building block G4"
 BASE COUNT
 14 a 3 c 8 g 7 t
 ORIGIN
 Query Match 100.0%; Score 32; DB 6; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.1;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 CAAGGTTGGAATTAAGAAATTAATTCGTGA 32
 Db 1 CAAGGTTGGAATTAAGAAATTAATTCGTGA 32
 RESULT 2
 OSJN00004/c 116498 bp DNA linear HTG 17-JUL-2002
 DEFINITION Oryza sativa chromosome 4 clone OSJNBA0014F04, *** SEQUENCING IN
 PROGRESS ***, 2 ordered pieces.
 AL606449
 VERSION AL606449.2 GI:21912451
 HTG: HTGS PHASE2
 Oryza sativa.
 Oryza sativa.
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryzae; Oryza.
 REFERENCE
 AUTHORS
 1 Fu, G., Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F.,
 Tu, Y. F., Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Kang, H.,
 Chen, X. Y., Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W.,
 Wang, L. J., Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L.,
 Zhu, F. H., Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q.,
 Hu, X., Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Meng, Q. J.,
 Zhang, L., Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. Q., Lu, Y.,
 Li, C., Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K.,
 Zhou, B., Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P.
 and Hong, G. P.
 Direct Submission
 Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,
 Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,
 CHINA. E-mail enquiries: dhan@ncgr.ac.cn. Clone requests:
 dhan@ncgr.ac.cn
 Oryza sativa Japonica (nipponbare) genomic DNA, chromosome 4, BAC
 clone: OSJNBA0014F04.
 On Jul 19, 2002 this sequence version replaced gi:15552687.
 Web site: http://www.ncgr.ac.cn
 ----- Summary Statistics
 Assembly program: phrap

 NOTE: This is a PHASE2 sequence. Gaps are shown by nnnn. Genes
 were identified by a combination of several methods: Gene
 prediction programs including Rgenesh (http://www.sotberry.com/),
 genescan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkES
 (http://genemark.biology.gatech.edu/GeneMark/), CRNAcan-SE (Sean
 Eddy, http://genome.muc1.edu/eddy/CRNAcan-SE/), searches of the
 complete sequence against NCBI non-redundant protein database (nr)
 (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 62574: contig of 62574 bp in length

* 62575 62774: gap of 200 bp
 * 62775 116498: contig of 53724 bp in length.
 FEATURES
 LOCATION/Qualifiers
 1. 116498
 /organism="Oryza sativa"
 /variety="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="4"
 /clone="OSJNBA0014F04"
 /clone_lib="CUGI-OSJNBA"
 BASE COUNT 32679 a 26432 c 26711 g 30475 t 201 others
 ORIGIN
 Query Match 70.0%; Score 22.4; DB 2; Length 116498;
 Best Local Similarity 81.2%; Pred. No. 1.5e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 1 CAAGGTTGGAATTAAGAAATTAATTCGTGA 32
 Db 13070 CAAGATTGCAATTTGGGAAATTAATTCGTGA 13039
 RESULT 3
 AC123085
 LOCUS AC123085
 DEFINITION Rattus norvegicus clone CH230-909, *** SEQUENCING IN PROGRESS ***,
 63 unordered pieces.
 AC123085
 VERSION AC123085.2 GI:21723033
 HTG: HTGS PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 167191)
 Alzubayr, D. M., Adams, C., Adio-Oduola, B., Ali-Adnan, F. R., Allen, C.,
 Albrooke, S. L., Amaral, H. C., Aze, J. R., Ayale, M., Banks, T.,
 Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonini, D.,
 Bouch, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N. P.,
 Buhay, C., Burich, P., Burkett, C., Butrell, K. L., Byrd, N. C.,
 Caron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R.,
 Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,
 Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,
 Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
 Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M.,
 Faller, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hollnabe, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Huylk, S., Hume, J., Jackson, L. E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C.,
 Kravovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
 Li, J., Li, Z., Lichtenage, O., Lien, C., Liu, J., Liu, W., Louieged, H.,
 Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawlin, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, E., Newkirk, S., Ogih, M., Okwou, G.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwou, G.,
 Ogunye, N., Oyler, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pichens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshari, N., Sleson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansy, J., Taylor, C., Taylor, T., Talford, B., Thomas, N., Thomas, S.,
 Umami, K., Vaequez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 167191)
Submitted (12-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 167191)
Morley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:21239777.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDXP
Center clone name: CH230-909
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114513 bases at least Q40
Consensus quality: 117131 bases at least Q30
Consensus quality: 119267 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1066: contig of 1066 bp in length
1067 1166: gap of unknown length
1167 2757: contig of 1591 bp in length
2758 2858: gap of unknown length
2859 4142: contig of 1285 bp in length
4143 4242: gap of unknown length
4243 5847: contig of 1605 bp in length
5848 5947: gap of unknown length
5948 7339: contig of 1392 bp in length
7340 7439: gap of unknown length
7440 8655: contig of 1216 bp in length
8656 8756: gap of unknown length
8757 10233: contig of 1478 bp in length
10234 10333: gap of unknown length
10334 12015: contig of 1682 bp in length
12016 12115: gap of unknown length
12116 13329: contig of 1114 bp in length
13330 13329: gap of unknown length
13331 15087: contig of 1758 bp in length
15088 15187: gap of unknown length
15188 16776: contig of 1589 bp in length
16777 16876: gap of unknown length
16878 18703: contig of 1827 bp in length
18704 18803: gap of unknown length
18804 20293: contig of 1490 bp in length
20294 20393: gap of unknown length
20394 21718: contig of 1325 bp in length
21719 21818: gap of unknown length

21819 23074: contig of 1256 bp in length
23075 23174: gap of unknown length
23175 24805: contig of 1631 bp in length
24806 24905: gap of unknown length
24906 26255: contig of 1350 bp in length
26256 26355: gap of unknown length
26356 27515: contig of 1160 bp in length
27516 28892: contig of 1277 bp in length
28893 28992: gap of unknown length
28993 30170: contig of 1178 bp in length
30171 30270: gap of unknown length
30271 31632: contig of 1362 bp in length
31633 31732: gap of unknown length
31733 33023: contig of 1291 bp in length
33024 33123: gap of unknown length
33124 34124: contig of 1001 bp in length
34125 34224: gap of unknown length
34225 35778: contig of 1554 bp in length
35779 35878: gap of unknown length
35879 37834: contig of 1956 bp in length
37835 37934: gap of unknown length
37935 40210: contig of 2276 bp in length
40211 40310: gap of unknown length
40311 41638: contig of 1328 bp in length
41639 41738: gap of unknown length
41739 43521: contig of 1783 bp in length
43522 43621: gap of unknown length
43622 45077: contig of 1456 bp in length
45078 45177: gap of unknown length
45178 46984: contig of 1807 bp in length
46985 47084: gap of unknown length
47085 48843: contig of 1759 bp in length
48844 48943: gap of unknown length
48944 50980: contig of 2037 bp in length
50981 51080: gap of unknown length
51081 52334: contig of 1254 bp in length
52335 52434: gap of unknown length
52435 54389: contig of 1955 bp in length
54390 54489: gap of unknown length
54490 56000: contig of 1511 bp in length
56001 56100: gap of unknown length
56101 58503: contig of 2403 bp in length
58504 58603: gap of unknown length
58604 60246: contig of 1643 bp in length
60247 60346: gap of unknown length
60347 60443: contig of 3087 bp in length
60444 63433: gap of unknown length
63434 63533: gap of unknown length
63534 66154: contig of 2621 bp in length
66155 66254: gap of unknown length
66255 67965: contig of 1711 bp in length
67966 68065: gap of unknown length
68066 70031: contig of 1966 bp in length
70032 70131: gap of unknown length
70132 72594: contig of 2463 bp in length
72595 72694: gap of unknown length
72695 75301: contig of 2607 bp in length
75302 75401: gap of unknown length
75402 79068: contig of 3667 bp in length
79069 79168: gap of unknown length
79169 80614: contig of 1446 bp in length
80615 80714: gap of unknown length
80715 83675: contig of 2961 bp in length
83676 83775: gap of unknown length
83776 86937: contig of 3162 bp in length
86938 87037: gap of unknown length
87039 90098: contig of 3061 bp in length
90099 90198: gap of unknown length
90199 92410: contig of 2212 bp in length
92411 92510: gap of unknown length
92511 94924: contig of 2414 bp in length
94925 95024: gap of unknown length
95025 100832: contig of 5808 bp in length

ACCESSION 270783
VERSION 270783.1 GI:1263486
KEYWORDS HTG; DNA directed RNA polymerase III; EF-hand; LIN-19; Spingomyelin phosphodiesterase; TAU; Transposase.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 9069613
PUBMED 9851916
REMARK The *C. elegans* Sequencing Consortium.
REFERENCE 2 (bases 1 to 41502)
AUTHORS Lloyd, C.R.
JOURNAL Direct Submission
TITLE Submitted (16-APR-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess@anger.ac.uk or rwenematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
EXCEPTIONS are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone ZK856. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone ZK856 is at 32458 in sequence Z78410.
The true right end of clone ZK856 is at 104 in sequence Z73969.
The start of this sequence (1..104) overlaps with the end of sequence Z78410.
The end of this sequence (41399..41502) overlaps with the start of sequence Z73969.
For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=ZK856>
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
location/Qualifiers
1..41502
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="ZK856"
join(2797..2859,2905..3039,3418..3519,3567..3771,3818..3927,3975..4182,4229..4359,4409..4705,4752,4995..5055,5102..5205,5370..5534,5637..6028,6074..6125)
/gene="ZK856.1"
join(2797..2859,2905..3039,3418..3519,3567..3771,3818..3927,3975..4182,4229..4359,4409..4705,4752,4995..5055,5102..5205,5370..5534,5637..6028,6074..6125)
/gene="ZK856.1"
/note="Similarity with the *C. elegans* LIN-19 protein, contains similarity to Pfam domain: PF00888 (Cullin family), Score=1185.1, E-value=0, N=1
CDNA EST YK5943.3 comes from this gene
CDNA EST YK5943.5 comes from this gene
CDNA EST YK16197.5 comes from this gene
CDNA EST YK16197.3 comes from this gene
CDNA EST YK462C3.3 comes from this gene
CDNA EST YK462C3.5 comes from this gene

CDNA EST YK310H9.3 comes from this gene
CDNA EST YK34897.3 comes from this gene
CDNA EST YK34897.5 comes from this gene
CDNA EST YK367F8.5 comes from this gene
CDNA EST YK539411.3 comes from this gene
CDNA EST YK570C12.3 comes from this gene
CDNA EST YK539411.5 comes from this gene
CDNA EST YK570C12.5 comes from this gene
/codon_start=1
/protein_id="CA94852.1"
/db_xref="GI:3881806"
/db_xref="SWISS-PROT:Q23639"
/translation="MQDEWSKADPIVHALHQKVTAPAMODLFYHYKITSWDDGPKIKRDIITRCINDYHANKRIRSLQDGSILYIKEMNFPYQOANIPFVKIDSSRRSVPEPEESIRVTMLEKNEIIFMNSISBOLVYALTYEERGNITDQNVIGIRSEFVALNDRAGEDPLVYROSPEEQIFQETBYVKIKGNILNIEGLVAVVYADKLEBEOQARARYLNNPSITGKMEKAVIALVASFETITLAECSKLASADVERLORLYRLIRRTSGIDITVLACITDTHIRITREGLANDRNNANENISTPERYVOQLMPDFCSLVREGFCDNARLTARDKAFRAVNDSDIFTEMMNKKGRITLSVESKCAELIYADQDLILRTQSLKLTSEIDEKLNQVLLVLYKVENKQVFMFRFAHLSRLILMSADQEKEMWVTLRREGCMPSDAVNKLRLQDIELKQNSSPFKALGTNNKNSIADSNMYLNGANGRGSSERRISLPREDPVPDEWPTGANGRPDLCTTPMAVLPFGNDRANDKISLETALATELPDAELNRTLSLVAYPKRYQILLCDVSTTYTARDFTDSYKFLNHDNPNVANGKSOQGRKVNLLIGRLQLSLEANAERHESIVALRELRYVEGLVSIILKRTYTLAQITMELVILKPLFIPNRKIIKEQIDWLIENKYMERRADINTVYVTS"
join(6516..6596,7028..7111,7164..7379,7429..7540,7584..7678)
/gene="ZK856.4"
join(6516..6596,7028..7111,7164..7379,7429..7540,7584..7678)
/gene="ZK856.4"
/codon_start=1
/protein_id="CA94851.1"
/db_xref="GI:3881805"
/translation="MFEHQLLKWKLYILLCILVASEALDFNPLHQLCWLQVAVILAKRTNSQAVAEIKNTCMNVLLDNNWVCNGLIEVSANFGWTPAVKLCNHLNMCASPSSDKRSKMSVYKGRPNELTKVLLDILNDEPVSQDFRQVAVNEALNPNNGNDICKMLQGHVTPNNNDIVHFLKKPQQAISVSGANE"
complement(join(7983..8215,8261..8426,8632..8932,8984..9175,9380..9520,9590..9715,9760..9852,9897..10179,10239..10410))
/gene="ZK856.3"
complement(join(7983..8215,8261..8426,8632..8932,8984..9175,9380..9520,9590..9715,9760..9852,9897..10179,10239..10410))
/gene="ZK856.3"
/note="Weak similarity with spingomyelin phosphodiesterase"
/codon_start=1
/protein_id="CA94850.1"
/db_xref="GI:3881804"
/translation="MNSFSANKAYVNFYFNKVPDADQFAFLKELSNAKCNPRISENCITSVHVAHIGPEVFEKTPNFMFRDENVERFLDLVRYANSIGMMIFGHHDTPLINDSKENNQALAMAPVTPWBSDLGACANNTPFRVETDASIKODISTYININDLNNKNSIPVFEISFKDAIGITGDINPTTMSALDKKNDITLFFKTIIDNTLNFMPQMPGAVYRGALQCSMEYSDYPRYSSCLSTYSALTRSLNLPFLFALITYELDLFSPSKVHLHFOLEFSDIHTRKNSYTHQPPKKKSPGNALILNLTGTFVNGMYMTCIPAPGIIKRAVOICPIDACQAVVQNDYERSARQVRLLELTGTFVNGMYMTCIPVKNHVAARTHEEREVEVPIGTTSRGTVLWLGIFASFTQONSVSQSTISFSTQSFSSYSLSRSRITSSSSQVABSCDSATITTPSPYPAQOSTGLNLPPTHTLPDPAFPRRTKIIPPEPMKQELAPVTOQIWNCTINSIRKGSDBIGRQSLSCSKMELBYNWTGSL"
complement(join(11155..12030,12072..12227))
/gene="ZK856.2"
complement(join(11155..12030,12072..12227))
/gene="ZK856.2"
/note="contains similarity to Pfam domain: PF01498 (Transposase), Score=208.2, E-value=4.1e-59, N=1"
/codon_start=1
/protein_id="CA94849.1"

```

/db_xref="GI:3881803"
/translacion="WKSVGCCNUSLDVKKAIIVAGFEGCIPTKMLALIOISPSPIYWK
VIKTQTEKSVALARISPGRPRVTFTSKMRNRARIEDPHRTADIDOMIISPNPVP
SIKTVRRRLQQAGLHGRKPVKPFITISKRNRAVRVMATAHLRMGDEWAKHIWSDESK
FNLFSGDSNSVVWRPVGSRSPKYOCPEYFKGGGVGMWGCTSSMGPLRIOSIMD
ROYENILETTTMTRPWALONGRGFPQODNDPKHTSLHSWFORRRHHLLDMWPSCIP
DLNPETHLMEELEERLRIGIRASNNADAKNOJLENMKAILPMSTHYTLIDSMPRCOAVI
DANGATATX")
complement(join(12949..13101,13147..13350))
/gene="ZK856.5"
complement(join(12949..13101,13147..13350))
/gene="ZK856.5"
/codon_start=1
/protein_id="CAA94853.2"
/db_xref="GI:18376575"
/db_xref="SPTREMBL:Q23640"
/translacion="MMPLMYLIIFLATLPILKSWGVOVLHADPHLDVYOYSINGDNKH
CHDSBERTLTIGEYGDCMDAPKRILLQLAIDESARIIPPNDLIWTBDNVAHIDGYG
WESEKFSCIYYOTLSII"
join(121936..21996,22068..22271,22325..22439,22490..22583)
22883..22945)
/gene="ZK856.6"
join(121936..21996,22068..22271,22325..22439,22490..22583)
22883..22945)
/gene="ZK856.6"
/codon_start=1
/protein_id="CA94854.1"
/db_xref="GI:3881808"
/db_xref="SPTREMBL:Q23641"
/translacion="MASVEPOLMTFLTIACQTVAUVUVEPIVESMSDEBAANVALDP
EPPVARSDPILIIRRKSIILVGLISEKTEVCCKTKIQHRHPQGHANSRVKEIODDES
SSLERTFECEGENRPPOHGVDHDLFIECVTVERPANVARKVSGBPYPATTIRVP
ILICEBRLROXPDKPEKKA"
join(125331..25464,25511..25742)
/gene="ZK856.7"
join(125331..25464,25511..25742)
/gene="ZK856.7"
/codon_start=1
/protein_id="CNA94855.1"
/db_xref="GI:3881809"
/db_xref="SPTREMBL:Q23642"
/translacion="MRVAVILIFFRSVALADIITYHLKSIDPDKRESGOIVASNKGKSE
RWSYHKLIKOKNTFOQKADECGKLKWELITTWDAGKPAIVAKTYLLDGIRVITYKGDLL
KAYQDRGAICIKGCAPL"
complement(join(25949..26187,26233..26457,26508..26564,
26642..26708))
/gene="ZK856.8"
complement(join(25949..26187,26233..26457,26508..26564,
26642..26708))
/gene="ZK856.8"
/notes="contains similarity to Pfam domain: PF00036 (EF

Query Match      56.9%; Score 21.4; DB 3; Length 41502;
Best Local Similarity    80.6%; Pred.No.4+02;
Matches     25; Conservative   0; Mismatches       6; Indels    0; Gaps    0;
```

REFERENCE
AUTHORS

Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
1 (bases 1 to 158886)
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
Bodet, B., Bock, J., Bowie, S., Brooks, A., Buhey, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Pashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraruto, D.,
Forcum-Tansey, J., Prant, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,
Holloway, C., Hoast, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondeljewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichterger, S., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
Oswal, G., Patis, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugeng, R.,
Tabors, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahab, M.,
Washington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,
Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D., and
Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Unpublished
2 (bases 1 to 158886)
Worley, K.C.

Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 26, 2000 this sequence version replaced gi:19719630.

COMMENT

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HMDP
Center clone name: RP11-70P20

Summary Statistics
Sequencing vector: M13; L08821
Assembly program: Phrap; version 0.990329
Library: Dye-terminator Big Dye; 100% of reads
Consensus quality: 144562 bases at least Q40
Consensus quality: 152026 bases at least Q30
Consensus quality: 154719 bases at least Q20
Estimated insert size: 15575; sum-of-contigs estimation
Estimated insert size: 209830; agarose-fp estimation
Quality coverage: 3.2x in Q20 bases; agarose-fp estimation
Quality coverage: 4.3x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
43752: contig of 43752 bp in length
43753: gap of unknown length
43853: gap of unknown length
71719: contig of 27867 bp in length
71720: gap of unknown length
71820: gap of 21548 bp in length
93368: gap of unknown length
93467: gap of unknown length
93468: contig of 15397 bp in length
108865: gap of unknown length
108866: gap of unknown length
124124: contig of 15160 bp in length
124125: gap of unknown length
124224: gap of unknown length
135573: contig of 11349 bp in length
135574: gap of unknown length
135673: gap of unknown length
141767: contig of 6094 bp in length

141768 141867: gap of unknown length
* 141868 149244: contig of 7377 bp in length
* 149245 149344: gap of unknown length
* 149345 155749: contig of 6405 bp in length
* 155750 155849: gap of unknown length
* 155850 157240: contig of 1391 bp in length
* 157241 157341 158886: contig of 1546 bp in length.
* 157341 158886: contig of 1546 bp in length.
Location/Qualifiers
1. 158886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-70P20"

BASE COUNT 45757 a 33003 c 33199 g 45917 t 1010 others
ORIGIN

Query Match 66.9%; Score 21.4; DB 2; Length 158886;
Best Local Similarity 80.6%; Pred. No. 3.4e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGATTAATTCGTGA 32
Db 105743 AAGGATTGAATTAAGGATTAATTCGTCA 105713

RESULT 8
AL161899 159981 bp DNA linear PRI 25-FEB-2001
LOCUS Human DNA sequence from clone RP11-418110 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL161899
VERSION AL161899.21 GI:13606073
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 159981)
Phillimore,B.
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12964402.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-418110 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-418110 The true
left end of clone RP11-347N5 is at 72478 in this sequence. The true

right end of clone RP11-307B20 is at 55792 in this sequence.
FEATURES
source
1. 159981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-418110"
/clone_lib="RPCT-11.2"
26800..26832
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest data."
36723..36895
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest."
37359..37715
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest."
40660..41106
/note="Single clone region. Sequence from reads from a
short insert library derived from a clone PCR. Restriction
digest data confirm the assembly."
67000..67044
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
67045..67080
/note="Sequence from overlapping clone AC021003 sequenced
by Whitehead Institute. Assembly confirmed by restriction
digest."
67081..67186
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
68186..68322
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
68323..68428
/note="Sequence from overlapping clone AC021003 sequenced
by Whitehead Institute. Assembly confirmed by restriction
digest."
68429..68534
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
76600..76875
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
92981..93194
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
96774..96941
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
99380..99401
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
124439..124492
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
128739..128745
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
130242..130333
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
153887..153985
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
30906 c 29579 g 47018 t

ORIGIN

Query Match 66.9%; Score 21.4; DB 9; Length 159981;
 Best Local Similarity 80.6%; Pred. No. 3.4e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
 |||||
 Db 148207 AAGGTTCTTAATGAGAGTAATTAATGTA 148237

RESULT 9
 AC025725
 LOCUS
 DEFINITION Caenorhabditis elegans clone Y69A2R, *** SEQUENCING IN PROGRESS
 AC025725 188215 bp DNA linear HTG 12-MAR-2000
 *** 1 unordered piece.

ACCESSION AC025725
 VERSION AC025725.1 GI:7230349
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 188215)
 Waterston, R.H.
 Unpublished
 The sequence of Caenorhabditis elegans clone
 2 (bases 1 to 188215)
 Waterston, R.H.
 Direct Submission
 Submitted (12-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 1 188215: contig of 188215 bp in length.
 1. 188215
 Location/Qualifiers
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y69A2R"

BASE COUNT 60803 a 33182 c 33097 g 61133 t

ORIGIN

Query Match 66.9%; Score 21.4; DB 2; Length 188215;
 Best Local Similarity 80.6%; Pred. No. 3.3e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
 |||||
 Db 10198 AAGCTTTGAAAAAAGAAATTAATTAATGTA 10228

RESULT 10
 AC084158
 LOCUS
 DEFINITION Caenorhabditis elegans cosmid Y69A2AR, complete sequence.
 AC084158 188616 bp DNA linear INV 19-APR-2002
 AC084158.1 GI:10800383
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 188616)
 Waterston, R.
 Unpublished
 The genome sequence of the nematode C. elegans: a platform for

JOURNAL
 MEDLINE
 PUBMED
 99069613
 9851916
 2 (bases 1 to 188616)
 Latreille, P.
 The sequence of C. elegans cosmid Y69A2AR
 Unpublished (2001)
 3 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (13-OCT-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 4 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (28-JUN-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (07-SEP-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (03-NOV-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:

COMMENT
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: tw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
 (www.wormbase.org/db/seq/sequence?name=Y69A2AR;class=Sequence)

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y69A2AL, 200 bp overlap; the 3' cosmid is Y94H6A,
 200 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/htmi/CE_INDEX.html) and The C.

JOURNAL Submitted (23-AUG-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 206320)
REFERENCE Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
AUTHORS Saepthimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
Haugen, E.D.
TITLE Direct Subassembly
JOURNAL Submitted (30-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Apr 30, 2002 this sequence version replaced gi:15281267.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UMG
Web site: <http://www.genome.washington.edu>
Contact: umgchgs@u.washington.edu
Drafting Center: BCM

----- Project Information

Center project name: chr-3
Center clone name: RP11-1035L22 (bc0587)

----- Summary Statistics

Sequencing vector: M13; 108821; 56% of reads
Chemistry: Dye-terminator; 44% of reads
Chemistry: Dye-terminator; 39% of reads
Chemistry: Dye-terminator; 7% of reads
Chemistry: Dye-terminator; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 206172 bases at least Q40
Consensus quality: 206299 bases at least Q30
Consensus quality: 206318 bases at least Q20
Insert size: 206320; sum-of-contigs
Quality coverage: 6.7x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5' Mapping in progress
3' RP11-70P20 (UMGC:bc0189) AC016944

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HandIII

BglII

NsiI

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

4134 4199 5750 5746 11730 10912

6382	6536	2067	2093	471	<800
512	<800	23113	23465	2365	2357
449	<800	2296	2352	1918	1892
2576	2435	2143	2093	3320	3336
763	<800	213	<800	3896	3912
14581	14639	1292	1270	805	801
511	<800	3330	3324	12	<800
1164	1137	5797	5746	3017	3120
233	<800	73	<800	2190	2175
5383	5502	4684	4673	1134	1110
2447	2435	3485	3483	7391	7328
5965	5912	6554	6644	2204	2175
3370	3384	6939	6843	698	<800
6272	6204	4859	4848	1913	1892
3968	3976	3635	3600	779	<800
2608	2583	6367	6354	2159	2175
2437	2435	113	<800	4727	4679
941	924	4916	4957	11125	10912
1001	997	2479	2581	948	953
4280	4139	16772	16122	522	<800
5303	5502	2964	2988	952	953
355	<800	6115	6171	4238	4193
4289	4497	2628	2767	11374	10912
2260	2251	4465	4493	863	888
2962	2900	395	<800	73	<800
3370	3384	2333	2352	5323	5658
7479	7313	5879	5914	405	<800
460	<800	389	<800	25	<800
1786	1820	4720	4673	1080	1110
9533	9486	372	<800	4217	4088
2233	2251	4480	4493	1913	1892
4983	5101	1783	1922	6083	5979
2490	2435	478	<800	3677	3912
263	<800	1532	1511	54	<800
1324	1369	7151	7155	906	888
218	<800	1634	1701	3963	3912
5060	5101	378	<800	872	888

3357	3384	4034	4034	1976	1892
7336	7313	1972	2093	256	<800
767	769	1729	1701	3436	3661
47	<800	1437	1511	1572	1541
2880	2900	685	<800	2872	2903
747	769	1545	1568	6731	6650
6644	6536	2758	2767	4845	4679
2866	2900	2779	2767	308	<800
1837	1820	1781	1701	3311	3326
5645	5502	313	<800	13161	12910
3776	3775	5859	5914	1924	1892
6322	6204	4313	4302	1681	1641
3825	3775	152	<800	5303	5230
1547	1651	1769	1701	881	888
1444	1369	1338	1405	11498	10912
2004	1967	1191	1152	2197	2175
1462	1492	2135	2093	23	<800

Query Match 66.9%; Score 21.4; DB 9; Length 206320;
Best Local Similarity 80.6%; Pred. No. 3.3e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTCA 32
Db 153840 AAGGATTGAATTAAGGAATTAATTCGTCA 153870

RESULT 12
AC006892/c
LOCUS
DEFINITION
AC006892
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

AC006892 299081 bp DNA linear HTG 26-FEB-1999
Caenorhabditis elegans clone Y69A2, *** SEQUENCING IN PROGRESS ***,
10 unordered pieces.
AC006892 GI:4309907
HTG; HTGS PHASE1.
Caenorhabditis elegans.
Caenorhabditis elegans.
Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 299081)
Waterston, R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 299081)
Waterston, R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence; version replaced gi:4263456.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3087: contig of 3087 bp in length
3088 3101: gap of unknown length
3102 7168: contig of 4067 bp in length
7169 7182: gap of unknown length
7183 16968: contig of 9786 bp in length
16969 16982: gap of unknown length
16983 30314: contig of 13311 bp in length
30314 30327: gap of unknown length
30328 45680: contig of 15353 bp in length
45681 45694: gap of unknown length
45695 77792: contig of 32098 bp in length
77793 77806: gap of unknown length
77807 122072: contig of 44266 bp in length
122073 122086: gap of unknown length
122087 157435: contig of 35349 bp in length
157436 157449: gap of unknown length
157450 216837: contig of 59388 bp in length
216838 216851: gap of unknown length
216852 299081: contig of 82230 bp in length.

Location/Qualifiers
1. 299081
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y69A2"

BASE COUNT 96473 a 52922 c 53657 g 95903 t 126 others
ORIGIN

Query Match 66.9%; Score 21.4; DB 2; Length 299081;
Best Local Similarity 80.6%; Pred. No. 3.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTCA 32
Db 35978 AAGGTTGGAATTAAGGAATTAATTCGTCA 35948

RESULT 13
AR016866/c
LOCUS
DEFINITION
AR016866
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

AR016866 1567 bp DNA linear PAT 05-DEC-1998
Sequence 100 from patent US 5777200.
AR016866 GI:3973143
AR016866.1 GI:3973143
Unknown.
Unknown.
Unclassified.
1 (bases 1 to 1567)
Ryals, J.A., Alexander, D.C., Goodman, R.M. and Stinson, J.R.
Chemically regulatable and anti-pathogenic DNA sequences and uses
thereof
Patent: US 5777200-A 100 07-JUL-1998;
Location/Qualifiers
1. 1567
/organism="unknown"
BASE COUNT 469 a 313 c 298 g 477 t 10 others
ORIGIN

Query Match 66.2%; Score 21.2; DB 6; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.1e+02;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATTAAGGAATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGGAATTAATTCGTCA 427

RESULT 14
AR020892/c
LOCUS
DEFINITION

AR020892 1567 bp DNA linear PAT 05-DEC-1998
Sequence 100 from patent US 5789214.

ACCESSION AR020892
 VERSION AR020892.1 GI:3975507
 KEYWORDS
 SOURCE Unknown.

ORGANISM Unknown.
 REFERENCE Unclassified.
 1 (bases 1 to 1567)
 AUTHORS Ryals,J.A., Friedrich, L.B., Uknes,S.J. and Ward,E.R.
 TITLE Method of inducing gene transcription in a plant
 JOURNAL Patent: US 5789214-A 100 04-AUG-1998;
 FEATURES Location/Qualifiers
 source 1..1567
 /organism="unknown"

BASE COUNT 469 a 313 c 298 g 477 t 10 others
 ORIGIN

Query Match 66.2%; Score 21.2; DB 6; Length 1567;
 Best Local Similarity 82.1%; Pred. No. 7.1e+02;
 Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTTCGAATATAAGGATTAATTCGTGA 32
 |||||
 Db 454 GGTTCGAATATAAGGATTAATTCGTGA 427

RESULT 15
 AR027215/c AR027215 1567 bp DNA linear PAT 29-SEP-1999
 LOCUS
 DEFINITION Sequence 100 from patent US 5856154.
 ACCESSION AR027215
 VERSION AR027215.1 GI:5938055
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE Unclassified.
 1 (bases 1 to 1567)
 AUTHORS Ryals,J.A., Alexander,D.C., Goodman,R.M. and Ward,E.R.
 TITLE Method of protecting plants from oomycete pathogens
 JOURNAL Patent: US 5856154-A 100 05-JAN-1999;
 FEATURES Location/Qualifiers
 source 1..1567
 /organism="unknown"

BASE COUNT 469 a 313 c 298 g 477 t 10 others
 ORIGIN

Query Match 66.2%; Score 21.2; DB 6; Length 1567;
 Best Local Similarity 82.1%; Pred. No. 7.1e+02;
 Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTTCGAATATAAGGATTAATTCGTGA 32
 |||||
 Db 454 GGTTCGAATATAAGGATTAATTCGTGA 427

Search completed: May 11, 2003, 06:19:46
 Job time : 527.927 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 62.2356 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-18
Perfect score: 32
Sequence: 1 caagggtcgaataaagaataattcgtga 32

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	32	21	AAA96852
2	21.2	66.2	1567	20	AAV62796
3	21.2	66.2	1567	20	AAV81680
4	21	65.6	2898	21	AAZ6404
5	20.8	65.0	3103	22	AAH54098
6	20.8	65.0	3438	22	AAH54877
7	20.8	65.0	3463	22	AAH54633
8	20.2	63.1	2894	23	ABL23374
9	20	62.5	360	22	AAF23810

10	20	62.5	1563	22	AAF23809	Escherichia coli s
11	19.8	61.9	49	22	AAH17286	Human ribosomal pr
12	19.8	61.9	241	23	ABV60512	Human prostate exp
13	19.8	61.9	468	22	ABA46292	Human breast cell
14	19.8	61.9	468	22	ABN56837	Human foetal liver
15	19.8	61.9	468	22	ABN26466	Human bone marrow
16	19.8	61.9	468	22	AAK04949	Human brain expres
17	19.8	61.9	468	22	AAK0480	Human bone marrow
18	19.8	61.9	468	22	AAI15092	Probe #5025 for ge
19	19.8	61.9	468	22	AAI156431	Probe #5117 used t
20	19.8	61.9	468	22	AAI04851	Probe #4842 used t
21	19.8	61.9	468	24	ABN05112	Human genome-deriv
22	19.8	61.9	850	22	ABN06688	Human CDNA clone (
23	19.8	61.9	4052	23	ABN09300	Drosophila melanog
24	19.2	60.0	466	22	AAK06757	Human brain expres
25	19.2	60.0	466	22	AAK32468	Human bone marrow
26	19.2	60.0	466	22	AAI38302	Probe #6988 used t
27	19.2	60.0	466	24	ABN07254	Human genome-deriv
28	19.2	60.0	3166	24	ABQ70738	Listeria monocytog
29	19.2	60.0	11866	5	AAH40141	Sequence of human
30	19.2	60.0	11873	5	AAH40176	Sequence of human
31	19.2	60.0	38059	22	AAH54018	Human factor IX (h
32	19.2	60.0	38059	24	ABN95627	Gene #2125 used to
33	19.2	60.0	267156	24	ABN68560	Kidney cancer rela
34	19.2	60.0	2944528	24	ABA03041	Listeria monocytog
35	19	59.4	1249	23	AAH72728	DNA encoding novel
36	19	59.4	6277	22	AAH46729	Tumour suppressor
37	19	59.4	6277	24	ABJ34041	Human immune syste
38	19	59.4	7630	24	ABJ33123	Human immune syste
39	19	59.4	11788	24	ABN99851	Rat pyruvate kinas
40	19	59.4	18434	24	ABJ34006	Human immune syste
41	19	59.4	36635	23	ABN07970	Drosophila melanog
42	19	59.4	40388	23	ABN26342	Drosophila melanog
43	19	59.4	94618	21	AAH22285	BAC containing rep
44	18.8	58.8	435	21	AAH79604	Pinus radiata cell
45	18.8	58.8	763	24	ABQ38656	Oligonucleotide fo

ALIGNMENTS

RESULT 1	
AAA96852	
ID	AAA96852 standard; DNA; 32 BP.
AC	AAA96852:
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Guide desoxynucleotide building block G4.
XX	
KW	Promoter, intergenic region; Commelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; ss.
XX	
OS	Synthetic.
XX	
PN	WO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
PF	29-MAR-2000; 2000WO-IB00370.
XX	
PR	29-MAR-1999; 99FR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
PI	Rance I, Gruber V, Theisen M;
XX	
DR	WPI; 2000-647238/62.
XX	
PT	Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced
 PT with sequence from promoter comprising green tissue expression region
 XX
 PS Disclosure; Page 24; 91pp; English.
 CC The present sequence represents a guide desoxynucleotide building
 CC block, which was used to link directional desoxynucleotide building
 CC blocks during construction of chimeric promoters of the invention.
 CC The specification describes chimeric expression promoters. These
 CC chimeric promoters comprise a nucleic acid sequence which is derived
 CC from a first plant promoter, in which a plant vascular expression
 CC promoter region is replaced with a nucleic acid sequence derived from
 CC a second plant promoter comprising a plant green tissue expression
 CC promoter region. Preferably, the first plant promoter originates from
 CC *Comelina yellow mottle virus*, and the second plant promoter originates
 CC from the *Cassava vein mosaic virus*. Especially, the promoters are
 CC derived from intergenic regions. The chimeric promoters are useful
 CC for producing transgenic plants.
 XX
 SQ Sequence 32 BP; 14 A; 3 C; 8 G; 7 T; 0 other;
 Query Match 100.0%; Score 32; DB 21; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.0023;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 CAGGGTTCGAATTAAGGAATTAATTCGTGA 32
 DB 1 CAGGGTTCGAATTAAGGAATTAATTCGTGA 32
 RESULT 2
 AAV62796/c
 ID AAV62796 standard; cDNA; 1567 BP.
 AC AAV62796;
 XX
 DT 05-MAR-1999 (first entry)
 XX
 DE Tobacco SAR CHX independent gene clone 11.3.8.
 XX
 KM Chemically regulatable DNA promoter; expression control; pesticide;
 KM herbicide tolerance; systemic acquired resistance gene; ss.
 XX
 OS *Nicotiana acuminata*.
 XX
 PN US5851766-A.
 XX
 PD 22-DEC-1998.
 XX
 PF 31-MAY-1995; 95US-0456262.
 XX
 PR 31-MAY-1995; 95US-0456262.
 XX
 PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Harms C, Ryals JA;
 XX
 DR WPI; 1999-080396/07.
 XX
 PT Isolating chemically regulatable DNA sequences in plants - useful
 PT for chemically controlling expression in transformed plants
 XX
 PS Example 40A; Column 255-258; 175pp; English.
 XX
 CC This sequence represents a clone of the tobacco systemic acquired
 CC resistance (SAR) cycloheximide (CHX) independent.
 CC This gene can be isolated using the method of the invention.
 CC The method is for isolating a chemically regulatable DNA promoter
 CC fragment from the 5' flanking region of a chemically regulatable gene in
 CC a plant tissue. The method allows isolation of sequences which will be
 CC useful for the controlled expression of genes, under the control of a
 CC non-coding regulatable sequence. This is useful in plants with a

CC herbicide or pesticide detoxification mechanism under the control of a
 CC chemical regulator; the regulator being applied before or with the
 CC herbicide or pesticide to give optimal tolerance. The promoter fragment
 CC is useful for controlling sequences which encode traits such as
 CC height, shape, development, male or female sterility, and the ability
 CC of the plant to withstand cold, heat, salt and drought. The chemical
 CC induction of the promoter allows the regulation of production of
 CC compounds, e.g. flavours, fragrances, pigments, natural sweeteners,
 CC industrial feedstocks, antimicrobials and pharmaceuticals, by
 CC biosynthesis or metabolic conversion, whose biosynthesis is controlled
 CC by endogenous or foreign genes. The method allows control over the time
 CC and rate of gene expression either throughout the whole plant, or in
 CC localized tissues, to achieve e.g. fungal or insect resistance by for
 CC instance ducting the leaves with the chemical regulator. Controlling the
 CC developmental processes by the application of a regulating chemical in
 CC e.g. the commercial production of cultivated crops allows processes such
 CC as germination, flower formation and fruit ripening to be synchronised at
 CC a given time.
 XX
 SQ Sequence 1567 BP; 469 A; 313 C; 298 G; 477 T; 10 other;
 Query Match 66.2%; Score 21.2; DB 20; Length 1567;
 Best Local Similarity 82.1%; Pred. No. 50;
 Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 OY 5 GGTTGGAATTAAGGAATTAATTCGTGA 32
 DB 454 GGTTGGAATTAAGGAATTAATTCATTA 427
 RESULT 3
 AAV81680/c
 ID AAV81680 standard; DNA; 1567 BP.
 AC AAV81680;
 XX
 DT 25-FEB-1999 (first entry)
 XX
 DE Tobacco protein-synthesis independent gene p11.3.8.
 XX
 KM Regulation; transcription; plant tissue; chimeric construction; PR;
 KM pathogenesis-related protein; anti-pathogenic; transgenic plant;
 KM beta-1,3-glucanase activity; pest resistance; ss.
 XX
 OS *Nicotiana sp.*
 XX
 PN US5847258-A.
 XX
 PD 08-DEC-1998.
 XX
 PF 31-MAY-1995; 95US-0457364.
 XX
 PR 31-MAY-1995; 95US-0457364.
 PR 08-MAR-1988; 88US-0165667.
 PR 06-FEB-1989; 89US-0305566.
 PR 24-MAR-1989; 89US-0329018.
 PR 20-JUN-1989; 89US-0368672.
 PR 20-OCT-1989; 89US-0425504.
 PR 07-SEP-1990; 90US-0580431.
 PR 21-DEC-1990; 90US-0632441.
 PR 01-APR-1991; 91US-0678378.
 PR 27-SEP-1991; 91US-0768122.
 PR 06-MAR-1992; 92US-0848506.
 PR 06-NOV-1992; 92US-0973197.
 PR 06-APR-1993; 93US-0042847.
 PR 12-APR-1993; 93US-0045957.
 PR 16-JUL-1993; 93US-0093301.
 PR 13-JAN-1994; 94US-0181271.
 XX
 PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Meyer MB, Payne GB, Ryals JA, Ward ER;

DR WPI; 1999-059180/05.
XX DNA encoding pathogenesis-related glucanase proteins - useful for
PT producing transgenic plants with enhanced disease or pest resistance
XX
PS Example 40; Column 249-252; 169pp; English.
XX
CC The present invention describes a DNA molecule encoding a
CC pathogenesis-related (PR) protein having beta-1,3-glucanase activity
CC selected from PR-2, PR-2', PR-2'', PR-N, PR-O and PR-O'. Also described
CC are: (i) a chimeric gene comprising the above DNA molecule linked to a
CC heterologous promoter; (ii) a vector containing the chimeric gene;
CC (iii) a host cell containing the chimeric gene; (iv) a transgenic plant
CC containing the chimeric gene; and (v) a seed from the transgenic plant.
CC The DNA molecule is used to produce transgenic plants with enhanced
CC disease or pest resistance. The present sequence represents a tobacco
CC protein-synthesis independent gene p11.3.8 from the present invention.
XX
SQ Sequence 1567 BP; 469 A; 313 C; 298 G; 477 T; 10 other;
Query Match 66.2%; Score 21.2; DB 20; Length 1567;
Best Local Similarity 82.1%; Pred. No. 50;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 5 GGTTGGAATTAAGGAATTAATTCGTGA 32
|||||
Db 454 GGTCTAATAATGAAGTAATTAATGATTA 427
RESULT 4
AAZ46404/C
ID AAZ46404 standard; DNA; 2898 BP.
XX
AC AAZ46404;
XX
DT 07-MAR-2000 (first entry)
XX
DE Penicillium funiculosum xylanase C DNA.
XX
KM Enzyme mixture; xylanase; beta-glucanase cellulase; feruloyl esterase;
KM digestibility; vegetable; cereal; livestock; supplement; nutrition;
KM food; energy; cell wall; phosphorus; usage; excretion; ammonia release;
KM battery; intensive; agriculture; ds.
XX
OS Penicillium funiculosum.
XX
FH Key Location/Qualifiers
FT 1..1588
FT /*cag= a
FT /number= 1
FT CDS 1317..2042
FT /*cag= b
FT /product= "Penicillium funiculosum xylanase C"
FT intron 1589..1642
FT /*cag= c
FT /number= 1
FT exon 1643..2898
FT /*cag= d
FT /number= 2
XX
XX WO957325-A2.
XX
XX 11-NOV-1999.
XX
XX PD 06-MAY-1999; 99WO-IB00856.
XX
XX PF 06-MAY-1999; 98EP-0401101.
XX
XX PR 06-MAY-1998; 98EP-0401101.
XX
XX PA (RHON) RHONE-POULENC ANIMAL NUTRITION SA.
XX
XX PA (RHOD) RHODIA CHIM.
XX
XX PI Sabatier A, Fish NM, Haigh NP;
XX

DR WPI; 2000-062159/05.
XX P-PSDB; AAY58149.
XX
PT New Penicillium funiculosum, useful for increasing the feed
PT digestibility for farm animals
XX
XX Claim 22; Fig 1; 68pp; English.
XX
CC This sequence represents Penicillium funiculosum genomic DNA encoding
CC xylanase C. The invention relates to a novel strain of Penicillium
CC funiculosum, deposited under the Budapest treaty as IMI No. 378536,
CC and an enzyme mixture obtainable from this strain, which includes
CC xylanases, beta-glucanase cellulases or feruloyl esterases. These
CC enzymes improve the digestibility of vegetable sources (particularly
CC cereals). Supplementations of the diets of pigs with xylanase increased
CC the energy digestibility by at least 6% as compared to unamplimented
CC diets, indicating that the enzyme is increasing the breakdown of
CC cell walls (particularly those of wheat), causing the release of
CC additional food energy in the small intestine. The enzyme mixture
CC from the novel Penicillium funiculosum strain is used for feeding
CC livestock, especially by improving the digestibility of cereals such
CC as wheat, rye, barley, triticale, oat, rice, oil seeds such as soya,
CC sunflower, rapeseed, and cereal by-products such as wheat bran. The
CC novel enzymes decrease phosphorus excretion, increase digestive use
CC of phosphorus, improve amino acid digestibility and reduce the ammonia
CC in the air of animal batteries used in intensive agriculture.
XX
SQ Sequence 2898 BP; 802 A; 705 C; 597 G; 794 T; 0 other;
Query Match 65.6%; Score 21; DB 21; Length 2898;
Best Local Similarity 82.8%; Pred. No. 62;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 4 GGTTGGAATTAAGGAATTAATTCGTGA 32
|||||
Db 2283 GGCTTAATAATGAATTAATTAATCTGA 2255
RESULT 5
AAH54098/C
ID AAH54098 standard; DNA; 3103 BP.
XX
AC AAH54098;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3462.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX PN 17-MAY-2001.
XX
XX PD 09-NOV-2000; 2000WO-US30782.
XX
XX PF 09-NOV-1999; 99US-0164258.
XX
XX PR 09-NOV-1999; 99US-0164258.
XX
XX PA (GLAX) GLAXO GROUP LTD.
XX
XX PI Kimmertly WJ;
XX
XX WPI; 2001-316495/33.
XX
DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1028-1030; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC

CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SO Sequence 3103 BP, 1088 A; 424 C; 535 G; 1056 T; 0 other;
Query Match 65.0%; Score 20.8; DB 22; Length 3103;
Best Local Similarly 78.1%; Pred. No. 75;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 1 CAAGGTTGGAATTAAGATAATTCGTGA 32
Db 71 CAATGTTGGAATTAATTAAGATAATTCGTGA 40
RESULT 6
AAH54877/c
ID AAH54877 standard; DNA; 3438 BP.
XX
AC AAH54877;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4421.
XX
KM *Staphylococcus epidermidis* SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS *Staphylococcus epidermidis*.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly MJ;
XX
PT WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1961-1962; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.

CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SO Sequence 3438 BP, 1130 A; 608 C; 547 G; 1153 T; 0 other;
Query Match 65.0%; Score 20.8; DB 22; Length 3438;
Best Local Similarly 78.1%; Pred. No. 75;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 1 CAAGGTTGGAATTAAGATAATTCGTGA 32
Db 1603 CAATGTTGGAATTAATTAAGATAATTCGTGA 1572
RESULT 7
AAH54633/c
ID AAH54633 standard; DNA; 3463 BP.
XX
AC AAH54633;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3997.
XX
KM *Staphylococcus epidermidis* SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS *Staphylococcus epidermidis*.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly MJ;
XX
PT WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1670-1671; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 3463 BP; 1157 A; 549 C; 641 G; 1116 T; 0 other;

Query Match 65.0%; Score 20.8; DB 22; Length 3463;
Best Local Similarity 78.1%; Pred. No. 76;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAAGGTTGGAATTAAGCAATTAATTCTGA 32
DB 2710 CAATTGTTGAAATTAATTAATAAGATTCGTGA 2679

RESULT 8
ABL23374/c
ID ABL23374 standard; DNA; 2894 BP.

XX ABL23374;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 21595.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200174042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-655660/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

XX Claim 1; SEQ ID NO 21595; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB157737-AB172072).

XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2894 BP; 698 A; 699 C; 752 G; 745 T; 0 other;

Query Match 63.1%; Score 20.2; DB 23; Length 2894;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GTTCGAAATTAAGCAATTAATTCTGT 30
|||||

DB 2568 GTTCGAAATTAATTAATTAATACGT 2544

RESULT 9

AAF23810

ID AAF23810 standard; DNA; 360 BP.

XX AAF23810;

XX 22-MAR-2001 (first entry)

XX Escherichia coli serotype-specific marker fragment.

XX Escherichia coli; serotype-specific marker; O157:H7; O157:NM; O55:H7;

XX serotype detection; ds.

XX Escherichia coli.

XX WO200077247-A1.

XX 21-DEC-2000.

XX 14-JUN-2000; 2000WO-CA00716.

XX 15-JUN-1999; 99US-0139260.

XX (UYGU-) UNIV GUELPH.

XX Chen S, Xu R, Li J;

XX WPI; 2001-112229/12.

XX New DNA markers useful for developing probes, primers and nucleic acid
XX based assays for detecting Escherichia coli serotypes O157:H7, O157:NM
XX and O55:H7 -

XX Example 1; Fig 2; 40pp; English.

XX The present sequence is a part of a novel DNA marker specific for
XX Escherichia coli serotypes O157:H7, O157:NM and O55:H7. The DNA marker is
XX useful for developing probes or primers that can be used to detect
XX E. coli serotypes O157:H7, O157:NM or O55:H7, as well as developing
XX nucleic acid based assays to detect E. coli serotypes in a biological
XX sample. The sample can be a clinical (e.g. blood, urine, serum, tears,
XX saliva, faeces or tissues), food, water or environmental sample.

XX Sequence 360 BP; 94 A; 91 C; 78 G; 97 T; 0 other;

Query Match 62.5%; Score 20; DB 22; Length 360;
Best Local Similarity 82.1%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGTTGCAATTAAGCAATTAATTCTGA 32
DB 159 GGTTGCAATTAAGCAATTAATTCTGA 186

RESULT 10

AAF23809

ID AAF23809 standard; DNA; 1583 BP.

XX AAF23809;

XX 22-MAR-2001 (first entry)

XX Escherichia coli serotype-specific 1583 bp marker.

XX Escherichia coli; serotype-specific marker; O157:H7; O157:NM; O55:H7;

XX serotype detection; ds.

XX Escherichia coli.

XX WO200077247-A1.

```

XX 21-DEC-2000.
PD
PF 14-JUN-2000; 2000WO-CA00716.
XX
PR 15-JUN-1999; 99US-0139260.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Chen S, Xu R, Li J;
XX
DR WPI; 2001-112229/12.
XX
PT New DNA markers useful for developing probes, primers and nucleic acid
PT based assays for detecting Escherichia coli serotypes O157:H7, O157:NM
PT and O55:H7 -
XX
XX
PS Claim 1; Fig 1; 40pp; English.
XX
CC The present sequence is a novel DNA marker specific for Escherichia coli
CC serotypes O157:H7, O157:NM and O55:H7. The DNA marker is useful for
CC developing probes or primers that can be used to detect E. coli serotypes
CC O157:H7, O157:NM or O55:H7, as well as developing nucleic acid based
CC assays to detect E. coli serotypes in a biological sample. The sample can
CC be a clinical (e.g. blood, urine, serum, tears, saliva, faeces or
CC tissues), food, water or environmental sample.
XX
SQ Sequence 1583 BP; 391 A; 430 C; 354 G; 408 T; 0 other;

Query Match 62.5%; Score 20; DB 22; Length 1583;
Best Local Similarity 82.1%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 GGTGGAATATAAGGAATTAATTCTGTA 32
Db 159 GGTGTGAATATGAGAAATTAATCTTGA 186

RESULT 11
ID AAD17286
XX AAD17286 standard; DNA; 49 BP.
XX
AC AAD17286;
XX
DT 29-NOV-2001 (first entry)
XX
DE Human ribosomal protein S24 probe #2 for MLPA assay.
XX
KW Human; multiplex ligation-dependent amplification; amplicon; probe;
KW single nucleotide polymorphism; ribosomal protein; S24; ss.
XX
OS Homo sapiens.
XX
PN WO200161033-A2.
XX
PD 23-AUG-2001.
XX
PF 15-FEB-2001; 2001WO-EP01739.
XX
PR 15-FEB-2000; 2000EP-0200506.
XX
XX
PA (SCHO/) SCHOUTEN J P.
XX
PI Schouten JP;
XX
DR WPI; 2001-550053/61.
XX
XX
PT An improved multiplex ligation-dependent amplification method for
PT detecting specific single stranded target nucleic acids in samples -
XX
XX Example 2; Page 151; 158pp; English.
XX
CC The invention relates to an improved multiplex ligation-dependent

```

```

CC amplification method for detecting specific single stranded target
CC nucleic acids in samples using a plurality of probe sets comprising at
CC least 2 probes. Each probe comprises a target specific region and a
CC non-complementary region comprising a primer binding site. The probes
CC in each set are ligated when hybridised to a target nucleic acid and
CC amplified by a primer set. The method is used for detecting a nucleotide
CC polymorphism, especially a single nucleotide polymorphism; detecting
CC multiple single stranded target nucleic acid sequences (through the
CC detection of multiple amplicons); determining the absolute or relative
CC abundance of multiple single stranded nucleic acids in a sample; and
CC detection of a break point region in rearranged nucleic acids. By using
CC a femtomolar amount of the probes, a large number of different probe
CC sets can be used to simultaneously detect and quantify a corresponding
CC large number of target sequences with high specificity. The present
CC DNA sequence is human ribosomal protein S24 probe which is used in a
CC multiplex ligatable probe amplification (MLPA) assay.
XX
SQ Sequence 49 BP; 19 A; 7 C; 14 G; 9 T; 0 other;

Query Match 61.9%; Score 19.8; DB 22; Length 49;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 AAGGTTGAAATTAAGGAATTAATTCTGTA 32
Db 10 AAGGTTGAAATTAAGACAGAAATTCGGA 40

RESULT 12
ABV60512
ID ABV60512 standard; cDNA; 241 BP.
XX
XX
AC ABV60512;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 60503.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JR;
XX
DR WPI; 2001-662795/76.
XX
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer; stage of prostate cancer -
XX
XX
PS Claim 1; Page 11524; 11750pp; English.
XX
XX
CC The invention relates to an isolated nucleic acid molecule (i) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (i) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;

```


CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 241 BP, 87 A, 37 C, 25 G, 92 T, 0 other;
Query Match 61.9%; Score 19.8; DB 23; Length 241;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 10 GAATTAAGGAATTAATTCTGTA 32
DB 98 GAAATTAAGGAATTAATTCTGTA 120
RESULT 13
ID ABA46292/c standard; DNA; 468 BP.
XX
XX ABA46292;
AC
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4987.
XX
XX Human; microarray; single exon probe; gene expression; breast;
KM disease; cancer; ss.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-496933/54.
DR
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 1; SEQ ID NO 4987; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical

CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarray. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 468 BP, 110 A, 126 C, 52 G, 180 T, 0 other;
Query Match 61.9%; Score 19.8; DB 22; Length 468;
Best Local Similarity 77.4%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 AAGGTTGGAATTAAGGAATTAATTCTGTA 32
DB 230 AATGAGAGGAATTAAGGAATTAAGGATGTA 200
RESULT 14
ID ABA56837/c standard; DNA; 468 BP.
XX
XX ABA56837;
AC
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #5142.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KM
XX Homo sapiens.
OS
XX WO200157277-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
PS Claim 1; SEQ ID NO 5142; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 468 BP, 110 A, 126 C, 52 G, 180 T, 0 other;
Query Match 61.9%; Score 19.8; DB 22; Length 468;
Best Local Similarity 77.4%; Pred. No. 1.6e+02;

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 12.8321 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32

Sequence: 1 caagggtcgaataaagaataatcgtga 32

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.2	66.2	1567	1 US-08-181-271A-100	Sequence 100, App
2	21.2	66.2	1567	1 US-08-449-315-100	Sequence 100, App
3	21.2	66.2	1567	1 US-08-444-803-100	Sequence 100, App
4	21.2	66.2	1567	1 US-08-449-043-100	Sequence 100, App
5	21.2	66.2	1567	1 US-08-456-265A-100	Sequence 100, App
6	21.2	66.2	1567	1 US-08-455-416-100	Sequence 100, App
7	21.2	66.2	1567	1 US-08-455-244-100	Sequence 100, App
8	21.2	66.2	1567	1 US-08-454-876-100	Sequence 100, App
9	21.2	66.2	1567	2 US-08-457-364-100	Sequence 100, App
10	21.2	66.2	1567	2 US-08-456-262-100	Sequence 100, App
11	21.2	66.2	1567	2 US-08-455-736-100	Sequence 100, App
12	21.2	66.2	1567	2 US-08-971-217-100	Sequence 100, App
13	21.2	66.2	1567	4 US-09-350-600-100	Sequence 100, App
14	21.2	66.2	1567	4 US-09-453-702B-242	Sequence 242, App
15	20	62.5	1683	4 US-09-009-443-6	Sequence 6, Appli
16	18.8	58.8	4307	3 US-08-944-594-1	Sequence 1, Appli
17	18.8	58.8	4307	3 US-08-103-663-35	Sequence 35, Appli
18	18.6	58.1	4550	4 US-08-194-180-2	Sequence 2, Appli
19	18.4	57.5	1560	1 US-08-973-462-2	Sequence 12, Appli
20	18.2	56.9	5361	4 US-08-973-462-1	Sequence 1, Appli
21	18.2	56.9	6152	4 US-08-641-638-412	Sequence 412, App
22	17.8	55.6	429	4 US-08-473-157A-3	Sequence 413, App
23	17.8	55.6	943	1 US-09-134-001C-2413	Sequence 2413, Ap
24	17.6	55.0	2049	4 US-09-155-078-1	Sequence 1, Appli
25	17.6	55.0	3431	4 US-09-155-078-1	Sequence 1, Appli
26	17.6	55.0	3431	4 US-09-155-078-1	Sequence 1, Appli
27	17.6	55.0	3431	4 US-09-155-078-1	Sequence 1, Appli

C	28	17.6	55.0	21338	4	US-08-961-527-20	Sequence 20, Appli
	29	17.4	54.4	2135	3	US-08-965-903B-7	Sequence 7, Appli
	30	17.4	54.4	2509	2	US-08-954-333-6	Sequence 6, Appli
	31	17.4	54.4	2797	2	US-08-954-333-8	Sequence 8, Appli
C	32	17.4	54.4	4155	1	US-07-876-280-1	Sequence 1, Appli
C	33	17.4	54.4	4155	1	US-07-875-772-1	Sequence 1, Appli
C	34	17.4	54.4	4155	1	US-08-063-170-1	Sequence 1, Appli
C	35	17.4	54.4	4155	1	US-08-158-232-1	Sequence 1, Appli
C	36	17.4	54.4	4155	1	US-08-304-626-1	Sequence 1, Appli
C	37	17.4	54.4	4155	1	US-08-316-301A-1	Sequence 1, Appli
C	38	17.4	54.4	4155	1	US-08-611-928-1	Sequence 1, Appli
C	39	17.4	54.4	4155	3	US-09-173-891-1	Sequence 1, Appli
C	40	17.4	54.4	4155	4	US-09-076-137-1	Sequence 1, Appli
C	41	17.4	54.4	4155	5	PCT-US92-03624-1	Sequence 1, Appli
C	42	17.4	54.4	4155	6	5281530-2	Sequence 1, Appli
C	43	17.4	54.4	4155	6	5426049-2	Sequence 1, Appli
C	44	17.4	54.4	9048	3	US-08-973-273-4	Sequence 4, Appli
C	45	17.2	53.8	1575	1	US-08-375-709-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-08-181-271A-100/c
Sequence 100, Application US/08181271A
Patent No. 5614395
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Umes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,271A
FILING DATE: 13-JAN-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-181-271A-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. NO. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCTTCGAATTAAGGATTAATTGCTGA 32
DB 454 GGTCTTAATTAAGGATTAATTGATTA 427

RESULT 2
US-08-449-315-100/c
Sequence 100, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesling, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.

APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Speirson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-449-315-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATTAAGCAATTAATTGCTGA 32
DB 454 GGTCTAAATTAAGCAATTAATTGATTA 427

RESULT 3
US-08-444-803-100/c
Sequence 100, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Ukenes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,803
FILING DATE: 19-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-444-803-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATTAAGCAATTAATTGCTGA 32
DB 454 GGTCTAAATTAAGCAATTAATTGATTA 427

RESULT 4
US-08-449-043-100/c
Sequence 100, Application US/08449043
Patent No. 5689044
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice

APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhau, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Utkes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericea C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/449,043
 FILING DATE: 24-MAY-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,432
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 100:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1567 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-449-043-100
 Query Match 66.2%; Score 21.2; DB 1; Length 1567;
 Best Local Similarity 82.1%; Pred. No. 7.4;
 Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 5 GGTTCGAATTAAGGATTAATTCGTGA 32
 DB 454 GGTTCGAATTAAGGATTAATTCGTGA 427
 RESULT 5
 US-08-456-265A-100/c
 Sequence 100, Application US/08456265A
 Patent No. 5767369
 GENERAL INFORMATION:
 APPLICANT: Alexander, Danny C.
 APPLICANT: Ryals, John A.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Stinson, Jeffrey R.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 520 White Plains Road, P.O. Box 2005
 CITY: Tarrytown
 STATE: New York
 COUNTRY: USA
 ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/456,265A
 FILING DATE: 31-MAY-95
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/181,271
 FILING DATE: 13-JAN-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727/DIV10
TELEPHONE: (919)541-8687
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
US-08-456-265A-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GGTTCGAATTAAGCAATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGCAATTAATTCGTCA 427

RESULT 6
US-08-455-416-100/c
Sequence 100, Application US/08455416
Patent No. 5777200
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moutoy, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Spertson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,416
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-455-416-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GGTTCGAATTAAGGATTAATTCGTGA 32
|||||
Db 454 GGTCTAATAATTAAGGATTAATTCGTGA 427

RESULT 7

US-08-455-244-100/c
Sequence 100, Application US/08455244
Patent No. 5789214

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaue, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,244
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-455-244-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GGTTCGAATTAAGGATTAATTCGTGA 32
|||||
Db 454 GGTCTAATAATTAAGGATTAATTCGTGA 427

RESULT 8

US-08-454-876-100/c
Sequence 100, Application US/08454876
Patent No. 5804693

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaue, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,876
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19925/PI/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-454-876-100
Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CY 5 GGTTCGAATTAAGGAAATTAATTCGTA 32
DB 454 GGTTCGAATTAAGGAAATTAATTCATTA 427
RESULT 9
US-08-457-364-100/c
Sequence 100, Application US/08457364
Patent No. 5847258
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,364
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 800
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847

;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/368,672
;; FILING DATE: 20-JUN-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8619
;; INFORMATION FOR SEQ ID NO: 100:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1567 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-457-364-100
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Cy 5 GATTGAATTAAGGATTAATTCGTGA 32
Db 454 GCTTCTAATAAAGGATTAATTCATTA 427
RESULT 10
US-08-456-262-100/C
; Sequence 100, Application US/08456262
; Patent No. 5851766
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedlich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Helms, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

;; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 106
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CIBA-GEIGY Corporation
;; STREET: 7 Skyline Drive
;; CITY: Hawthorne
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10532
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/456,262
;; FILING DATE: 31-MAY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/181,271
;; FILING DATE: 13-JAN-94
;; APPLICATION NUMBER: US 08/093,301
;; FILING DATE: 16-JUL-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/937,197
;; FILING DATE: 6-NOV-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/678,378
;; FILING DATE: 1-APR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/305,566
;; FILING DATE: 6-FEB-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/165,667
;; FILING DATE: 8-MAR-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/042,847
;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/368,672
;; FILING DATE: 20-JUN-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8619
;; INFORMATION FOR SEQ ID NO: 100:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1567 base pairs
;; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-456-262-100

Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTGGAATAAGGAATTAATTCGTA 32
Db 454 GGTCTAATAATTAAGGAATTAATTAATTA 427

RESULT 11
US-08-456-240-100/c
Sequence 100, Application US/08456240
Patent No. 5856154
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,240
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-456-240-100
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTGGAATAAGGAATTAATTCGTA 32
Db 454 GGTCTAATAATTAAGGAATTAATTAATTA 427
RESULT 12
US-08-455-736-100/c
Sequence 100, Application US/08455736
Patent No. 5880328
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,736
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-1994
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-455-736-100
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. NO. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTTCGAATTAAGGATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGGATTAATTCGTCA 427
RESULT 13
US-08-971-217-100/C
Sequence 100, Application US/08971217
Patent No. 5942662
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Harms, Christian
APPLICANT: Friedrich, Leslie
APPLICANT: Beck, James
APPLICANT: Uknes, Scott
APPLICANT: Ward, Eric
TITLE OF INVENTION: INDUCIBLE HERBICIDE RESISTANCE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5942662artis Corporation
STREET: 3054 Cornwallis Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,217
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504

FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727/DIV5/CONF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-971-217-100
MOLECULE TYPE: DNA (genomic)

Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATAAGGATTAATTCGTCA 32
DB 454 GGTTCGAATAAGGATTAATTCGTCA 427

RESULT 14
US-09-350-600-100/c
Sequence 100, Application US/09350600
Patent No. 6262342
GENERAL INFORMATION:
APPLICANT: Meigs, Frederick
APPLICANT: Shinshi, Hideaki
APPLICANT: Wenzler, Herman
APPLICANT: Hofsteenge, Jan
APPLICANT: Ryals, John
APPLICANT: Sperisen, Christoph
TITLE OF INVENTION: DNA SEQUENCES ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING BETA-1,3-GLUCANASE ACTIVITY
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6262342artis Corporation
STREET: 3054 Cornwallis Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,600
FILING DATE:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/971,217
FILING DATE: 14-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/381,443
FILING DATE: 18-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/353,312
FILING DATE: 17-MAY-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/226,303
FILING DATE: 29-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-198250
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-350-600-100

Query Match 66.2%; Score 21.2; DB 4; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTTCGAATTAAGGAATTAATTCGTGA 32
|||||
Db 454 GGTTCGAATTAAGGAATTAATTCGTGA 427

RESULT 15

US-09-453-702B-242/c
Sequence 242. Application US/09453702B
Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
Perna, Nicole T.
Burland, Valerie
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pluckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 242:

SEQUENCE CHARACTERISTICS:

LENGTH: 31880

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 242:

US-09-453-702B-242

Query Match 62.5%; Score 20; DB 4; Length 31880;

Best Local Similarity 82.1%; Pred. No. 30;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 GGTTCGAATTAAGGAATTAATTCGTGA 32
|||||
Db 28206 GGTTCGAATTAAGGAATTAATTCGTGA 28179

Search completed: May 11, 2003, 03:08:10
Job time : 17.8321 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 514.406 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32
Sequence: 1 caagggctcgaataaagaataaattcgtga 32

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_esthu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gse:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pla:*
21: em_gse_vrt:*
22: em_gse_fun:*
23: em_gse_man:*
24: em_gse_mus:*
25: em_gse_other:*
26: em_gse_pro:*
27: em_gse_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	71.9	728	17	AQ449597 500002C12
C 2	22.4	70.0	144	9	AV174753 AV174753
C 3	21.6	67.5	1016	17	BH147730 BH147730
C 4	21.4	66.9	309	10	BB490355 BB490355
C 5	21.4	66.9	375	17	AZ478586 AZ478586
C 6	21.4	66.9	432	13	BI327328 BI327328

Result No.	Score	Query Match	Length	ID	Description
C 7	21	65.6	1074	17	CNS06XG0
C 8	20.8	65.0	365	17	AQ100322
C 9	20.8	65.0	502	17	AQ214572
C 10	20.8	65.0	546	13	BI746889
C 11	20.8	65.0	560	17	AQ237213
C 12	20.8	65.0	662	17	AQ239775
C 13	20.8	65.0	669	17	CNS045XL
C 14	20.8	65.0	692	17	AG144326
C 15	20.8	65.0	768	10	AV853564
C 16	20.8	65.0	900	17	AQ746633
C 17	20.6	64.4	422	17	AQ734580
C 18	20.6	64.4	467	10	BE235639
C 19	20.6	64.4	532	13	BI742345
C 20	20.6	64.4	573	10	BE235635
C 21	20.6	64.4	1160	17	AZ936007
C 22	20.4	63.7	445	17	AQ664107
C 23	20.4	63.7	449	10	AM864304
C 24	20.4	63.7	479	10	BB829456
C 25	20.4	63.7	484	10	BB829743
C 26	20.4	63.7	535	17	AQ429861
C 27	20.4	63.7	639	12	BE885951
C 28	20.4	63.7	1066	17	CNS02GEH
C 29	20.2	63.1	218	9	AV078828
C 30	20.2	63.1	630	17	AZ198427
C 31	20.2	63.1	697	17	AZ360780
C 32	20.2	63.1	847	12	BE866488
C 33	20.2	63.1	971	17	CNS07DG5
C 34	20.2	63.1	1031	17	CNS00JB7
C 35	20.2	62.5	145	17	BH408228
C 36	20.2	62.5	296	9	A1059555
C 37	20.2	62.5	507	12	BF549568
C 38	20.2	62.5	531	10	BE082568
C 39	20.2	62.5	551	17	AQ838015
C 40	20.2	62.5	606	17	BH766421
C 41	20.2	62.5	607	17	AZ612822
C 42	20.2	62.5	776	13	BI656752
C 43	20.2	62.5	803	9	AL663807
C 44	20.2	62.5	867	12	BF675988
C 45	20.2	62.5	1226	14	BQ677337

ALIGNMENTS

RESULT 1
LOCUS AQ449597 728 bp DNA linear GSS 08-APR-1999
DEFINITION 500002C12.x2 CpIOWAM13mp18gDNA1 Cryptosporidium parvum genomic, DNA
ACCESSION AQ449597
VERSION AQ449597.1 GI:4578734
KEYWORDS GSS.
SOURCE Cryptosporidium parvum.
ORGANISM Cryptosporidium parvum
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeritida;
Cryptosporidiidae; Cryptosporidium.
REFERENCE 1 (bases 1 to 728)
AUTHORS Hyman,R.W., Fung,E., Qin,F., Rowley,D. and Davis,R.W.
TITLE Cryptosporidium parvum genome sequencing demonstration project
JOURNAL Unpublished (1999)
COMMENT Contact: Hyman, R. W.
Stanford DNA Sequencing and Technology Center
Stanford University School of Medicine, Palo Alto
855 California Avenue, Palo Alto, CA 94304, USA
Tel: 650 812 1972
Fax: 650 812 1975
Email: hyman@sequence.stanford.edu
For Annotation Data see http://medefgh.ucsf.edu/id/CpTags/home.html
Seq primer: M13(-21) forward
Class: shotgun.
Location/Qualifiers
1..728
/organism="Cryptosporidium parvum"

```

/db_xref="toma"  
//_clone_1ib="CpIONA1m1mp18gDNA1"  
/_lab_host="E. coli DH125  
/_note="vector: M13mp18; Site_1: Hind III; C. parvum (IOMA isolate) genomic DNA was hydrodynamically sheared to produce fragments having a tight size distribution between 1.5 and 3 kb. Adaptors (pgTAC/CAT/CAAGCGCTGAGp) were ligated to the randomly sheared gDNA fragments and pACTGTTC linker were ligated to the Hind III-cleaved M13mp18 vector. The adaptor-containing inserts were annealed and ligated to the vector and transformed into E. coli strain DH125. Recombinant phagemid clones from the first plating of the library were randomly selected for sequence analysis using the M13(-21) forward primer."
BASE COUNT      *   176 a     133 c       91 g        328 t
ORIGIN
Query Match          71.9%; Score 23; DB 17; Length 728;  
Best Local Similarity 83.9%; Pred. NO. 1.1e+03;  
Matches    26; Conservative    0; Mismatches    5; Indels    0; Gaps    0;
Oy         2 AAGGGTTCCAAATAAGAATTAATTCTGGA 32  
|||  
Df         680 AATGTTGCAAATTAAGCAAABAAATTCTGGA 650  
  
RESULT 2  
AV174753/c      144 bp      mRNA      linear EST 07-JUL-1999  
DEFINITION AV174753 Mus musculus C57BL/6J 8-day embryo Mus musculus cDNA clone  
LOCUS 5730412A1S, mRNA sequence.  
ACCSSION AV174753  
VERSION AV174753.1 GI:5398849  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Muskulus  
Bukaryota Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.  
REFERENCE 1 (bases 1 to 144)  
AUTHORS Carninci P., Shibata K., Ozawa Y., Komori H., Itoh M., Atzawa K., Akahira S., Akiyama J., Fukuda S., Fukuishi Y., Funayama T., Hara A., Hayatsu N., Horii F., Ishikawa T., Ito M., Izawa M., Kawai U., Kituchi N., Kojima Y., Matsuyama T., Mitsuma H., Oda H., Owa C., Sato K., Shihata Y., Shimamoto Y., Shiraki T., Sugabe Y., Suganaga Y., Suzuki H., Tateo M., Tomaru Y., Tomimaga N., Watanabe S., Yamagami M., Yanamura T., Yokota T., Yoshino M., Miyamatsu M., Okazaki Y. and Hayashizaki Y.  
RIKEN Mouse ESTs  
Unpublished (1999)  
Contact: Chile Owa  
Genome Science Laboratory  
RIKEN  
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan  
Tel.: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-research@riken.go.jp  
Thermotransformation and thermoinactivation of thermostable enzymes by chitosane and its application for the synthesis of full length CDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome.rtc.riken-go.jp) for further details.
```

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dT) primer [5' GTGATCCCATCTCGTAAGTGAGACGCCGCAGATGCTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bernaldo."

BASE COUNT 32 a 39 c 22 g 51 t

ORIGIN

Query Match 70.0%; Score 22.4; DB 9; Length 144;
Best Local Similarity 81.2%; Pred. No. 3e+03;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

CY 1 CAAGGGTCGAATAAAGCAATTAATTCGTGA 32
 ||||| ||||| ||||| |||||
Db 114 CAAGGGGTATTAAATATGAAACACTTCGTGA 83

RESULT 3
LOCUS BH147730 1016 bp DNA linear GSS 27-AUG-2001
DEFINITION ENT0822R Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic C, DNA sequence.
ACCESSION BH147730
VERSION BH147730.1 GI:15305567
KEYWORDS GSS.

SOURCE
ORGANISM Entamoeba histolytica.
Entamoeba histolytica
Eukaryota; Eumetazoa; Excavata; Amoebozoa; Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 1016)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
Unpublished (2001)

REFERENCE
AUTHORS Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: ML3-Reverse
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 492.
Location/Qualifiers

FEATURES
SOURCE

1 . 1016
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site 1: Not I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Bartell, Oxford University Press, 1999)."

BASE COUNT 387 a 147 c 175 g 307 t

Query Match 67.5%; Score 21.6; DB 17; Length 1016;
 Best Local Similarity 85.7%; Pred. No. 2.5e+03;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAGGGTTGGAATTAAGGAAATTAATTC 28
 Db 962 CAGGGTTGGAGAAAAAGATAATTC 989

RESULT 4
 BB490355/c 309 bp mRNA linear EST 25-JUL-2000
 LOCUS BB490355 RIKEN full-length enriched, 13 days embryo stomach Mus
 DEFINITION musculus cDNA clone D530014M20 3', mRNA sequence.
 ACCESSION BB490355
 VERSION BB490355.1 GI:9448982
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 309)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, U., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, P., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iwawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Salto, H., Sakai, C., Salto, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, P., Tomihata, N., Toya, T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamashita, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Yoshinide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermolabile and thermoactivated enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, U., Shibata, K., Iwawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source Location/Qualifiers

1. 309
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D530014M20"
 /clone_1ib="RIKEN full-length enriched, 13 days embryo stomach"
 /issue_type="stomach"
 /dev_stage="13 days embryo"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia"

BASE COUNT
 ORIGIN 83 a 80 c 57 g 88 t 1 others

Query Match 66.9%; Score 21.4; DB 10; Length 309;
 Best Local Similarity 80.6%; Pred. No. 4.4e+03;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CAAGGGTTGGAATTAAGGAAATTAATTCG 31
 Db 61 CAGGGTTGGAATTAAGGAAATTAATTCG 31

RESULT 5
 AZ478586/c 375 bp DNA linear GSS 04-OCT-2000
 LOCUS AZ478586
 DEFINITION 1M0298G17R Mouse 10kb plasmid UGICM library Mus musculus genomic
 clone UGICM0298G17 R. DNA sequence.
 ACCESSION AZ478586
 VERSION AZ478586.1 GI:10637592
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 375)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00
 Plate: 0298 row: G column: 17
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 high quality sequence stop: 375.
 Location/Qualifiers

FEATURES
 source

1. 375
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGICM0298G17"
 /clone_1ib="Mouse 10kb plasmid UGICM library"
 /sex="Male"
 /lab_host="B. Coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: PMD42ny. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1473214[g1473214]pAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 129 a 77 c 66 g 103 t

Query Match 66.9%; Score 21.4; DB 17; Length 375;
Best Local Similarity 80.6%; Pred. No. 4.1e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CAAGGTTGGAATTAAGAAATTAATTCGTG 31
Db 356 CAAGTGTGGAATTAAGAAATTAAGGTGTG 326

RESULT 6
BI327328/c 432 bp mRNA linear EST 01-SEP-2001
LOCUS AR077H097HYBH095 Porcine Thymus cDNA library Sus scrofa cDNA, mRNA
DEFINITION
SEQUENCE
VERSION BI327328
KEYWORDS BI327328.1 GI:15417031
SOURCE EST.
ORGANISM pig.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
TITLE 1 (bases 1 to 432)
Rink A., Santeschi, E.M. and Beattie, C.M.
Amplified, Normalized cDNA Libraries from a Porcine Model of Orthopedic Implant Associated Staphylococcus aureus Infection Unpublished (2001)

JOURNAL
COMMENT Contact: Rink A.
Department of Animal Biotechnology
College of Agriculture, Biotechnology and Natural Resources,
University of Nevada, Reno
MS 202, RA 103, 1664 N Virginia St, Reno, NV 89557-0236, USA
Tel: 775 784 1705
Fax: 775 784 1375

FEATURES
SOURCE Email: arink@cabnr.unr.edu
Tissues and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of Staphylococcus
aureus in a tibial transection, reduced and internally fixed with a
dynamic compression plate. NOTE: The sequences contain a cDNA
adapter between the EcoRI site and the start of the EST. The
adapter sequence is 'AATTCGCGACGAG'.

Location/Qualifiers
1..432
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone_id="Porcine Thymus cDNA library"
/issue_type="Thymus"
/cell_type="mixed"
/dev_stage="control, 5 month old castrated male"
/lab_host="SOLR"
/note="Vector: pSK, Site 1: Eco RI; Site 2: XhoI; Tissues
and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of
Staphylococcus aureus in a tibial transection, reduced and
internally fixed with a dynamic compression plate. NOTE:
The sequences contain a cDNA adapter between the EcoRI
site and the start of the EST. The adapter sequence is

BASE COUNT 148 a 97 c 78 g 100 t 9 others

Query Match 66.9%; Score 21.4; DB 13; Length 432;
Best Local Similarity 80.6%; Pred. No. 3.9e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 AAGGTTGGAATTAAGAAATTAATTCGTGA 32
Db 393 AAGATTGGAATTAAGAAATTAATTCGGA 363

RESULT 7
CNS06XG0 1074 bp DNA linear GSS 06-JUL-2001
LOCUS T3 end of clone XAX0AA002D12 of library XAX0AA from strain CBS 7064
DEFINITION
SEQUENCE
VERSION AL419686
KEYWORDS AL419686.1 GI:12202864
SOURCE GSS.
ORGANISM Pichia farinosa.

REFERENCE
AUTHORS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 1074)
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brotier, P., Cararegola, S.,
de Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuvéglise, C., Ozier-Kalogiropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nicche, C., Weslowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)

JOURNAL
MEDLINE 20584711
TITLE 2 (bases 1 to 1074)
de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Potier, S.
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
sorbitophila

JOURNAL
MEDLINE 20584725
TITLE 3 (bases 1 to 1074)
FEMS Lett. 487 (1), 87-90 (2000)

REFERENCE
AUTHORS Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Varrovia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
SOURCE Location/Qualifiers
1..1074
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone_id="XAX0AA002D12"
/clone_lib="XAX0AA"
/note="end : T3"

BASE COUNT 324 a 213 c 218 g 316 t 3 others

Query Match 65.6%; Score 21; DB 17; Length 1074;

Best Local Similarity 82.8%; Pred. No. 3.7e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAGGTTTCGAAATTAAGGATAATTCG 29
Db 920 CAAGGCTTAAGAAATTAAGGATAATTCG 948

RESULT 8
LOCUS A0100322 365 bp DNA linear GSS 27-AUG-1998

DEFINITION HS-3054_A1.D06.MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3054 Col=11 Row=G, DNA sequence.

ACCESSION A0100322
VERSION A0100322.1 GI:3471351

KEYWORDS GSS.
SOURCE human.

ORGANISM

REFERENCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL Mahairas GG, Wallace JC, Hood L
MEDLINE High Throughput Sequencing Center
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3054 row: G column: 11
Class: BAC ends
High quality sequence stop: 365.

FEATURES
source location/Qualifiers

1..365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3054 Col=11 Row=G"
/clone_1b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 114 a 67 c 74 g 110 t

ORIGIN

Query Match 65.0%; Score 20.8; DB 17; Length 365;
Best Local Similarity 78.1%; Pred. No. 6.3e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAGGTTTCGAAATTAAGGATAATTCGTA 32
Db 83 CAAGGTTTCGAAATTAAGGATAATTCGTA 114

RESULT 9
LOCUS A0214572 502 bp DNA linear GSS 18-SEP-1998

DEFINITION HS-3318_B2.G02.T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3218 Col=4 Row=N, DNA sequence.

ACCESSION A0214572
VERSION A0214572.1 GI:3625773

KEYWORDS GSS.
SOURCE human.

ORGANISM

REFERENCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 502)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL Mahairas GG, Wallace JC, Hood L
MEDLINE High Throughput Sequencing Center
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3218 row: N column: 4
Class: BAC ends
High quality sequence stop: 502.

FEATURES
source location/Qualifiers

1..502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3218 Col=4 Row=N"
/clone_1b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 147 a 91 c 105 g 158 t 1 others

ORIGIN

Query Match 65.0%; Score 20.8; DB 17; Length 502;
Best Local Similarity 78.1%; Pred. No. 5.6e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAGGTTTCGAAATTAAGGATAATTCGTA 32
Db 209 CAAGGTTTCGAAATTAAGGATAATTCGTA 240

RESULT 10
LOCUS B1746889 546 bp mRNA linear EST 25-SEP-2001

DEFINITION tm33c01.y1 Meloidogyme arenaria egg pMMP1 v1 Chiapelli McCarter
B1746889
B1746889
B1746889.1 GI:15768691

ACCESSION B1746889
VERSION B1746889.1 GI:15768691

KEYWORDS Meloidogyme arenaria.
SOURCE Meloidogyme arenaria.
ORGANISM Meloidogyme arenaria.
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heterodermidae; Meloidogyninae; Meloidogyme.

REFERENCE 1 (bases 1 to 546)
McCarter,J., Clifton,S., Chiapelli,B., Pepe,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,
Rinko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steepce
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.

The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wustl.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapelli@wustl.wustl.edu & jmcarter@wustl.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North Carolina

State University.
Seq primer: -40RP from GIBCO
High quality sequence stop: 439.

FEATURES

Source

1.546
/organism="Meloidogyne arenaria"
/db_xref="taxon:6304"
/clone_lib="Meloidogyne arenaria egg PAMP1 v1 Chiapelli
McCarter"
/dev_store="egg"
/lab_host="DH10B"
/note="Vector: PAMP1 (GIBCO); Site_1: NotI; Site_2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dt priming (Dyna).
PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Directionally
cloned into the UDG sites of PAMP1. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North
Carolina State University."

BASE COUNT

245 a 52 c 75 g 174 t

ORIGIN

Query Match 65.0%; Score 20.8; DB 13; Length 546;
Best Local Similarity 78.1%; Pred. No. 5.4e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAGGTTGCAATTAAGATTAATTCGTGA 32

Db 77 CAAGGTTGCAATTAAGATTAATTCATTA 108

RESULT 11

LOCUS

AQ237213 560 bp DNA linear GSS 21-APR-1999
DEFINITION RPII1-65J23.TJ RPII-11 Homo sapiens genomic clone RPII-11-65J23,
DNA sequence.

ACCESSION

AQ237213
AQ237213.1 GI:3669504

VERSION

GSS.

KEYWORDS

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

Adams, M.D., Rounsley, S.D., Zhao, S., Baas, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

AUTHORS

Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)

TITLE

Other GSSs: RPII1-65J23.TK

JOURNAL

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

COMMENT

Tel: 301 838 0200
Fax: 301 838 0208

E-mail: mdadams@igf.org

Clones are derived from the human BAC library RPII-11. For BAC
library availability, please contact Pieter de Jong
(pdejong@igf.org, med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1.560

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752481.4"
/db_xref="taxon:9606"
/clone="RPII-11-65J23"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

ORIGIN

1.560

BASE COUNT

245 a 52 c 75 g 174 t

ORIGIN

1.560

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752481.4"
/db_xref="taxon:9606"
/clone="RPII-11-65J23"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPII1 Human Male BAC Library"

BASE COUNT

192 a 72 c 130 g 165 t

ORIGIN

Query Match 65.0%; Score 20.8; DB 17; Length 560;
Best Local Similarity 78.1%; Pred. No. 5.4e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CAAGGTTGCAATTAAGATTAATTCGTGA 32

Db 370 CAAGGTTGCAATTAAGATTAATTCCTTA 401

RESULT 12

LOCUS

AQ239775 662 bp DNA linear GSS 21-APR-1999
DEFINITION RPII1-64O15.TK RPII-11 Homo sapiens genomic clone RPII-11-64O15,
DNA sequence.

ACCESSION

AQ239775
AQ239775.1 GI:3668527

VERSION

GSS.

KEYWORDS

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

Adams, M.D., Rounsley, S.D., Zhao, S., Baas, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

AUTHORS

Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)

TITLE

Other GSSs: RPII1-64O15.TJ

JOURNAL

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

COMMENT

Tel: 301 838 0200
Fax: 301 838 0208

E-mail: mdadams@igf.org

Clones are derived from the human BAC library RPII-11. For BAC
library availability, please contact Pieter de Jong
(pdejong@igf.org, med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1.662

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752482.2"
/db_xref="taxon:9606"
/clone="RPII-11-64O15"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

ORIGIN

220 a 109 c 122 g 211 t

BASE COUNT

220 a 109 c 122 g 211 t

ORIGIN

220 a 109 c 122 g 211 t

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752482.2"
/db_xref="taxon:9606"
/clone="RPII-11-64O15"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

ORIGIN

220 a 109 c 122 g 211 t

BASE COUNT

220 a 109 c 122 g 211 t

ORIGIN

220 a 109 c 122 g 211 t

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752482.2"
/db_xref="taxon:9606"
/clone="RPII-11-64O15"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

ORIGIN

220 a 109 c 122 g 211 t

BASE COUNT

220 a 109 c 122 g 211 t

ORIGIN

220 a 109 c 122 g 211 t

FEATURES

/organism="Homo sapiens"
/db_xref="GDB:752482.2"
/db_xref="taxon:9606"
/clone="RPII-11-64O15"
/clone_lib="RPII-11"
/sex="Male"

Source

Location/Qualifiers

LOCUS CNS045XL 669 bp DNA linear GSS 18-MAY-2000
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-Ori end of clone
 084H09 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL275970.1 GI:8010173
 VERSION GSS; genome survey sequence.
 KEYWORDS Tetradon nigroviridis.
 SOURCE Tetradon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetradon.
 1 (bases 1 to 669)
 Reest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished
 2 (bases 1 to 669)
 Reest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 3 (bases 1 to 669)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.
 Location/Qualifiers
 1..669
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="084H09"
 /clone_1lb="G"
 /note="Genoscope sequence ID : COBG084CD05SP1-end ;
 PUC-Ori"
 BASE COUNT 156 a 159 c 138 g 210 t 6 others
 ORIGIN
 Query Match 65.0%; Score 20.8; DB 17; Length 669;
 Best Local Similarity 78.1%; Pred. No. 5e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 CAAGGTTCCGAATTAAGGATTAATTCGTGA 32
 Db 217 CAATGGTTCTAAGGAGGAAATTAATTCCTGA 186
 RESULT 14
 AGI44326 692 bp DNA linear GSS 08-JAN-2002
 LOCUS AGI44326
 DEFINITION Pan troglodytes DNA, clone: RP43-005E15.TJ, genomic survey
 sequence.
 ACCESSION AGI44326
 VERSION AGI44326.1 GI:16674004
 KEYWORDS GSS.
 SOURCE Pan troglodytes male lymphocytes DNA, clone_1lb:RP43-43 Chimpanzee
 Male BAC Library clone:RP43-005E15.TJ.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 BAC end sequences of library RP43-43
 Unpublished

REFERENCE 2 (bases 1 to 692)
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimbases@sc.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library RP43-43 This BAC
 end was generated during the RAD process and may have higher chance
 of clone tracking errors.
 PRIMERS
 Sequencing: TJ
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI.
 R.Site 2 : EcoRI.
 Location/Qualifiers
 1..692
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-005E15.TJ"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_1lb="RP43-43 Chimpanzee Male BAC Library"
 BASE COUNT 229 a 133 c 125 g 205 t
 ORIGIN
 Query Match 65.0%; Score 20.8; DB 17; Length 692;
 Best Local Similarity 78.1%; Pred. No. 5e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 CAAGGTTCCGAATTAAGGATTAATTCGTGA 32
 Db 198 CAAGGTTCTTAAGGAGGAAATTAATTCCTTA 229
 RESULT 15
 AV853564/c 768 bp mRNA linear EST 08-NOV-2001
 LOCUS AV853564
 DEFINITION AV853564 Nori Satoh unpublished cDNA library, larva Clona
 intestinalis cDNA clone rcilv13h17 3', mRNA sequence.
 ACCESSION AV853564
 VERSION AV853564.1 GI:166839832
 KEYWORDS EST.
 SOURCE Clona intestinalis.
 ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Clona.
 1 (bases 1 to 768)
 Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
 Expressed genes in Clona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 Location/Qualifiers
 1..768
 /organism="Clona intestinalis"
 /db_xref="taxon:7719"
 /clone="rcilv13h17"
 /clone_1lb="Nori Satoh unpublished cDNA library, larva"
 /tissue_type="whole animal"
 /dev_stage="larva"
 /note="Vector: pBluescript SK"
 BASE COUNT 212 a 138 c 174 g 242 t 2 others
 ORIGIN

Query Match 65.0%; Score 20.8; DB 10; Length 768;
 Best Local Similarity 78.1%; Pred. No. 4.8e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CAAGGTTGAAATTAAGGAATTAATTGCTGA 32
 |||||
 Db 382 CCAGAGTTCGGAATTAATCGAATTAATGCTGA 351

Search completed: May 11, 2003, 04:54:38
 Job time : 518.406 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: May 11, 2003, 03:08:16 ; Search time 31.6792 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32
Sequence: 1 caagggttcgaataaagaataatcgtga 32

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	32	9 US-09-963-803-18	Sequence 18, Appl
2	20	62.5	31880	9 US-10-114-170-242	Sequence 242, App
3	19.8	61.9	468	10 US-09-864-761-4932	Sequence 4932, App
4	19.4	60.6	590	10 US-09-864-761-14873	Sequence 14873, A
5	19.2	60.0	466	10 US-09-864-761-11326	Sequence 11326, A
6	19.2	60.0	38059	10 US-09-880-107-2125	Sequence 2125, App
7	19.2	60.0	684973	10 US-09-263-959-1	Sequence 1, Appl1
8	18.8	58.8	435	9 US-10-101-464A-405	Sequence 405, Appl
9	18.8	58.8	9834	7 US-08-781-986A-37	Sequence 37, Appl
10	18.6	58.1	1524	10 US-09-815-242-6547	Sequence 6547, App
11	18.6	58.1	2328	9 US-09-938-842A-2679	Sequence 2679, App
12	18.6	58.1	32768	10 US-09-070-927A-128	Sequence 128, App
13	18.4	57.5	121	9 US-10-092-154-1884	Sequence 1884, App
14	18.4	57.5	121	10 US-09-764-847-1884	Sequence 1884, App
15	18.2	56.9	184	10 US-09-864-761-26009	Sequence 26009, App
16	18.2	56.9	566	10 US-09-864-761-9558	Sequence 9558, App
17	18.2	56.9	597	10 US-09-864-761-13097	Sequence 13097, App
18	18.2	56.9	764	10 US-09-770-445-942	Sequence 942, App
19	18.2	56.9	796	10 US-09-770-445-829	Sequence 829, App

20	18.2	56.9	1238	9 US-10-174-363-31	Sequence 31, Appl
21	18.2	56.9	1545	10 US-09-815-242-4389	Sequence 4389, App
22	18.2	56.9	1725	10 US-09-815-242-8228	Sequence 8228, App
23	18.2	56.9	1861	10 US-09-822-830A-4	Sequence 4, Appl1
24	18.2	56.9	2102	10 US-09-764-864-268	Sequence 268, App
25	18.2	56.9	5361	9 US-09-742-096-2	Sequence 2, Appl1
26	18.2	56.9	6152	9 US-09-742-096-1	Sequence 1, Appl1
27	18.2	56.9	8593	9 US-09-764-904-135	Sequence 135, App
28	18.2	56.9	8593	9 US-10-091-548-135	Sequence 135, App
29	18.2	56.9	8593	9 US-10-074-095-1196	Sequence 1196, App
30	18.2	56.9	8593	10 US-09-764-860-1196	Sequence 1196, App
31	18.2	56.9	335913	9 US-09-754-853A-2	Sequence 2, Appl1
32	18.2	56.9	335913	9 US-09-754-853A-3	Sequence 3, Appl1
33	18.2	56.2	383	10 US-09-770-791-130	Sequence 130, App
34	18.2	56.2	433	9 US-09-918-995-8370	Sequence 8370, App
35	18.2	56.2	491	9 US-09-918-995-8666	Sequence 8666, App
36	18.2	56.2	614	10 US-09-770-149-843	Sequence 843, App
37	18.2	56.2	1020	9 US-10-121-857-69	Sequence 69, Appl
38	18.2	56.2	2000	9 US-09-938-842A-4262	Sequence 4262, App
39	18.2	56.2	11360	9 US-09-764-891-6508	Sequence 6508, App
40	18.2	56.2	11360	9 US-09-764-891-7415	Sequence 7415, App
41	18.2	56.2	55795	10 US-09-880-107-1543	Sequence 1543, App
42	18.2	56.2	640681	10 US-09-790-988-1	Sequence 1, Appl1
43	17.8	55.6	284	10 US-09-560-863-726	Sequence 726, App
44	17.8	55.6	327	10 US-09-974-300-3882	Sequence 3882, App
45	17.8	55.6	406	9 US-09-918-995-4876	Sequence 4876, App

ALIGNMENTS

```

RESULT 1
US-09-963-803-18
Sequence 18, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 18
LENGTH: 32
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Guide desoxynucleotide building
OTHER INFORMATION: block G4
US-09-963-803-18

Query Match      100.0%; Score 32; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAGGTTTCGAATAAAGATAAATTCGTGA 32
Db 1 CAAGGTTTCGAATAAAGATAAATTCGTGA 32

RESULT 2
US-10-114-170-242/c
Sequence 242, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blactner, Frederick R.
Burland, Valerie

```

```

Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 242:
SEQUENCE CHARACTERISTICS:
LENGTH: 31880
TYPE: nucleic acid
STRANDEDNESS: double
* TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 242:
US-10-114-170-242
Query Match 62.5%; Score 20; DB 9; Length 31880;
Best Local Similarity 82.1%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 5 GGTGGAATTAAGGAATTAATTCGTGA 32
Db 28206 GGTGTGAAATATGGAATTAATTTCTGA 28179
RESULT 3
US-09-864-761-4932/c
Sequence 4932, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/633,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
```

```

PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4932
LENGTH: 468
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031679.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-4932
Query Match 61.9%; Score 19.8; DB 10; Length 468;
Best Local Similarity 77.4%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
Db 230 AATCGAGAGAAATTAAGGAATTAATGCTGA 200
RESULT 4
US-09-864-761-14873
Sequence 14873, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
```



```
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14873
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL157883.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
US-09-864-761-14873
```

```
Query Match 60.6%; Score 19.4; DB 10; Length 590;
Best Local Similarity 79.3%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
OY 2 AAGGTTGGAATTAAGGATTAATTCT 30
DB 343 AAGATTGAAAAAAGATTAAATTCAT 371
```

RESULT 5
US-09-864-761-11326/C

```
; Sequence 11326, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11326
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005774.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.65
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.59
US-09-864-761-11326
```

```
Query Match 60.0%; Score 19.2; DB 10; Length 466;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
OY 1 CAAGGTTGGAATTAAGGATTAATTCTGA 32
DB 188 CAAGGTTGTGATTAAGGATTAATTTCTTA 157
```

RESULT 6

```
; Sequence 2125, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
```

NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2125
LENGTH: 38059
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 K02402
US-09-880-107-2125

Query Match 60.0%; Score 19.2; DB 10; Length 38059;
Best Local Similarity 75.0%; Pred. No. 6.4e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAGGCTCGAATTAAGCAATTAATTCGTGA 32
DB 27113 CAAGGCTCGAATTAAGCAATTAATTCGTGA 271144

RESULT 7
US-09-263-959-1

Sequence 1, Application US/09263959
Patent No. US20020150891A1
GENERAL INFORMATION:
APPLICANT: Hood, Leroy E.
APPLICANT: Rowen, Lee
APPLICANT: Koop, Ben F.
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
NUMBER OF SEQUENCES: 1279
CORRESPONDENCE ADDRESS:
ADDRESSER: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,959
FILING DATE: 05-MAR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 920010.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 684973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-263-959-1

Query Match 60.0%; Score 19.2; DB 10; Length 684973;
Best Local Similarity 75.0%; Pred. No. 8.4e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAGGCTCGAATTAAGCAATTAATTCGTGA 32
DB 224824 CAAGGCTTCGATTAAGCAATTAATTCCTTA 224855

RESULT 8
US-10-101-464A-405/C
Sequence 405, Application US/10101464A
Publication No. US20030046728A1

GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signaling
FILE REFERENCE: 11000.1020c2
CURRENT APPLICATION NUMBER: US/10/101,464A
CURRENT FILING DATE: 2002-03-18
PRIOR APPLICATION NUMBER: 09/704,302
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 09/228,986
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/162,866
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: PCT/US00/00724
PRIOR FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 989
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 405
LENGTH: 435
TYPE: DNA
ORGANISM: Pinus radiata
US-10-101-464A-405

Query Match 58.8%; Score 18.8; DB 9; Length 435;
Best Local Similarity 76.7%; Pred. No. 3.8e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAAGGCTTCGAATTAAGCAATTAATTCGT 30
DB 313 CAAGGATTCGCCAAGAATAATTAATTCCT 284

RESULT 9
US-08-781-986A-37

Sequence 37, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 9834 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-781-986A-37

Query Match
Best Local Similarity 58.8%; Score 18.8; DB 7; Length 9834;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 AGGCTCGAATAAAGATTAATTGCTGA 32
Db 5950 AGCTTAGTAAGAAGACTAATTGCTGA 5979

RESULT 10
US-09-815-242-6547
Sequence 6547, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PatsSeq for Windows Version 4.0
SEQ ID NO 6547
LENGTH: 1524
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1524)
US-09-815-242-6547

Query Match
Best Local Similarity 58.1%; Score 18.6; DB 10; Length 1524;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TTCGAATTAAGATAAATTGCTG 31
Db 472 TACGAATTAATGAAGAAATTCGTG 496

RESULT 11
US-09-938-842A-2679
Sequence 2679, Application US/09938842A
Patent No. US20020160378A1

GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2679
LENGTH: 2328
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2679

Query Match
Best Local Similarity 58.1%; Score 18.6; DB 9; Length 2328;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAGGTTGCAATAAAGATTAATA 25
Db 1217 CAAGAGTTGCAAGAAGAGATTAATA 1241

RESULT 12
US-09-070-927A-128
Sequence 128, Application US/09070927A
Patent No. US20020120116A1

GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 128;
US-09-070-927A-128

Query Match 58.1%; Score 18.6; DB 10; Length 32768;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TTCGAAATTAAGAAATTAATTCGTG 31
DB 19349 TACGAAATTAATGAAGAAATTCGTG 19373

RESULT 13

US-10-092-154-1884/C
; Sequence 1884, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092.154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1884
; LENGTH: 121
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1884

Query Match 57.5%; Score 18.4; DB 9; Length 121;
Best Local Similarity 78.6%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGGTTGAAATTAAGAAATTAATTCGT 30
DB 115 ATGGTTCAAAATTAAGGAGTAATTCGT 88

RESULT 14

US-09-764-847-1884/C
; Sequence 1884, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764.847
; CURRENT FILING DATE: 2001-01-17
; Prior Application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1884
; LENGTH: 121
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1884

Query Match 57.5%; Score 18.4; DB 10; Length 121;
Best Local Similarity 78.6%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGGTTGAAATTAAGAAATTAATTCGT 30
DB 115 ATGGTTCAAAATTAAGGAGTAATTCGT 88

RESULT 15

US-09-864-761-26009
; Sequence 26009, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Neomica-X-1
CURRENT APPLICATION NUMBER: US/09/864.761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26009
LENGTH: 184
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO AC008166.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.3
OTHER INFORMATION: EST HUMAN HIT: BE077941.1, EVALUATE 7.00e-04
OTHER INFORMATION: SWISSPROT HIT: P10895, EVALUATE 1.00e+00
OTHER INFORMATION: NT HIT: AL163268.2, EVALUATE 7.00e-03
US-09-864-761-26009

Query Match 56.9%; Score 18.2; DB 10; Length 184;
Best Local Similarity 74.2%; Pred. No. 5.5e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AAGGTTGAAATTAAGAAATTAATTCGTA 32
DB 143 ATGGATTGAAAGAAAGAAATTAATTCGTA 173

Mon May 12 14:15:16 2003

us-09-963-803-18.rnpb

Page 7

Search completed: May 11, 2003, 06:32:41
Job time : 110.679 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 688.446 Seconds
(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctcagactatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

- 1: gb_ba:*
- 2: gb_hg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_ov:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pl:*
- 10: gb_ro:*
- 11: gb_ro:*
- 12: gb_ey:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_ba:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_mu:*
- 21: em_mu:*
- 22: em_mu:*
- 23: em_mu:*
- 24: em_mu:*
- 25: em_mu:*
- 26: em_mu:*
- 27: em_mu:*
- 28: em_mu:*
- 29: em_mu:*
- 30: em_hg_hum:*
- 31: em_hg_hum:*
- 32: em_hg_hum:*
- 33: em_hg_hum:*
- 34: em_hg_hum:*
- 35: em_hg_hum:*
- 36: em_hg_hum:*
- 37: em_hg_hum:*
- 38: em_hg_hum:*
- 39: em_hg_hum:*
- 40: em_hg_hum:*
- 41: em_hg_hum:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	65	AX036742	AX036742 Sequence
2	65	100.0	317	AX036737	AX036737 Sequence
3	65	100.0	371	AX036739	AX036739 Sequence
4	58	89.2	348	AX036738	AX036738 Sequence
5	58	89.2	398	AX036740	AX036740 Sequence
6	58	89.2	472	AX036759	AX036759 Sequence
7	58	89.2	541	AX036758	AX036758 Sequence
8	58	89.2	604	AX036757	AX036757 Sequence
9	55	84.6	332	AX036755	AX036755 Sequence
10	55	84.6	393	AX036753	AX036753 Sequence
11	55	84.6	462	AX036754	AX036754 Sequence
12	55	84.6	600	AX036756	AX036756 Sequence
13	55	84.6	9285	AX036757	AX036757 Sequence
14	55	84.6	15077	AX036758	AX036758 Sequence
15	47	72.3	7489	AX036759	AX036759 Sequence
16	46	70.8	243	AX036735	AX036735 Sequence
17	28.8	44.3	1381	AX036737	AX036737 Sequence
18	28.8	43.4	81480	AX036738	AX036738 Sequence
19	27.8	42.8	168367	AX036739	AX036739 Sequence
20	27.4	42.2	214175	AX036740	AX036740 Sequence
21	27.2	41.8	16357	AX036753	AX036753 Sequence
22	27.2	41.8	169862	AX036754	AX036754 Sequence
23	27.2	41.8	172346	AX036755	AX036755 Sequence
24	26.8	41.2	142559	AX036756	AX036756 Sequence
25	26.8	41.2	145535	AX036757	AX036757 Sequence
26	26.8	41.2	148792	AX036758	AX036758 Sequence
27	26.8	41.2	149055	AX036759	AX036759 Sequence
28	26.8	41.2	164564	AX036735	AX036735 Sequence
29	26.8	41.2	178814	AX036737	AX036737 Sequence
30	26.8	41.2	186491	AX036738	AX036738 Sequence
31	26.8	41.2	195240	AX036739	AX036739 Sequence
32	26.8	41.2	196877	AX036740	AX036740 Sequence
33	26.8	41.2	242569	AX036753	AX036753 Sequence
34	26.8	40.9	128250	AX036754	AX036754 Sequence
35	26.6	40.9	146024	AX036755	AX036755 Sequence
36	26.6	40.9	177514	AX036756	AX036756 Sequence
37	26.6	40.9	194418	AX036757	AX036757 Sequence
38	26.4	40.6	50979	AX036758	AX036758 Sequence
39	26.4	40.6	164719	AX036759	AX036759 Sequence
40	26.4	40.6	335871	AX036735	AX036735 Sequence
41	26.2	40.3	636	AX036737	AX036737 Sequence
42	26.2	40.3	156764	AX036738	AX036738 Sequence
43	26	40.0	2566	AX036739	AX036739 Sequence
44	26	40.0	104383	AX036740	AX036740 Sequence
45	26	40.0	138556	AX036753	AX036753 Sequence

ALIGNMENTS

RESULT 1
AX036742
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX036742
Sequence 8 from Patent WO0058485.
AX036742
AX036742.1 GI:11226251
synthetic construct.
synthetic construct.
artificial sequence.
1 (bases 1 to 65)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 8 05-OCT-2000;
JOURNAL

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. 65
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S1"

BASE COUNT 20 a 16 c 14 g 15 t
ORIGIN

Query Match 100.0%; Score 65; DB 6; Length 65;
Best Local Similarity 100.0%; Pred. No. 5.2e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 60
Db 1 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 60

Qy 61 TAGCT 65
Db 61 TAGCT 65

RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.

REFERENCE
1 (bases 1 to 317)

AUTHORS
TITLE
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 3 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES
SOURCE
1. 317
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1116"

Promoter 1. 317
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN

Query Match 100.0%; Score 65; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 60
Db 8 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 67

Qy 61 TAGCT 65
Db 68 TAGCT 72

RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.

REFERENCE
1 (bases 1 to 371)

AUTHORS
TITLE
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 5 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES
SOURCE
1. 371
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1116"

Promoter 1. 371
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN

Query Match 100.0%; Score 65; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 6.2e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 60
Db 8 CATGTCGAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAA 67

Qy 61 TAGCT 65
Db 68 TAGCT 72

RESULT 4
AX036738 348 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 4 from Patent WO0058485.
ACCESSION AX036738
VERSION AX036738.1 GI:11226247
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.

REFERENCE
1 (bases 1 to 348)

AUTHORS
TITLE
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 4 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES
SOURCE
1. 348
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1117"

Promoter 1. 348
BASE COUNT 116 a 70 c 78 g 84 t
ORIGIN

Query Match 89.2%; Score 58; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAAAGCT 65
Db 49 CAGACTAGTATCCGCGTCATCATGACATCATCATCAGTACTGAGAGATGAAAGCT 106

RESULT 5
AX036740 398 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 6 from Patent WO0058485.
ACCESSION AX036740
VERSION AX036740.1 GI:11226249
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.

REFERENCE
1 (bases 1 to 398)

REFERENCE 1 (bases 1 to 398)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 6 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 398
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1147"
BASE COUNT promoter 128 a 80 c 93 g 97 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 6
LOCUS AX036759 472 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 25 from Patent WO0058485.
ACCESSION AX036759
VERSION AX036759.1 GI:11226268
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE 1 (bases 1 to 472)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 472
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1169"
BASE COUNT promoter 149 a 92 c 112 g 119 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 7
LOCUS AX036758 541 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 24 from Patent WO0058485.
ACCESSION AX036758
VERSION AX036758.1 GI:11226267
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE 1 (bases 1 to 541)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow

JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 24 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 541
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1168"
BASE COUNT promoter 169 a 104 c 130 g 138 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 541;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 8
LOCUS AX036757 604 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 23 from Patent WO0058485.
ACCESSION AX036757
VERSION AX036757.1 GI:11226266
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE 1 (bases 1 to 604)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 23 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 604
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1167"
BASE COUNT promoter 186 a 116 c 145 g 157 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 604;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 9
LOCUS AX036755 392 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE 1 (bases 1 to 392)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES ; GRUBER VERONIQUE (FR)
source 1. .392
/db_xref="taxon:32630"
/note="promoter MPr1164"

Promoter 127 a 80 c 87 g 98 t
BASE COUNT
ORIGIN

Query Match 84.6%; Score 55; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
Db 20 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 10
AX036753 393 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 19 from Patent WO058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 393)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 19 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES Location/Qualifiers
source 1. .393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1162"

Promoter 128 a 75 c 93 g 97 t
BASE COUNT
ORIGIN

Query Match 84.6%; Score 55; DB 6; Length 393;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
Db 20 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 11
AX036754 462 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 20 from Patent WO058485.
ACCESSION AX036754
VERSION AX036754.1 GI:11226263
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 462)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 20 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES Location/Qualifiers
source 1. .462

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1163"

Promoter 148 a 87 c 111 g 116 t
BASE COUNT
ORIGIN

Query Match 84.6%; Score 55; DB 6; Length 462;
Best Local Similarity 100.0%; Pred. No. 6.5e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
Db 20 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 12
AX036756 600 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 22 from Patent WO058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 600)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 22 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES Location/Qualifiers
source 1. .600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1165"

Promoter 188 a 111 c 147 g 154 t
BASE COUNT
ORIGIN

Query Match 84.6%; Score 55; DB 6; Length 600;
Best Local Similarity 100.0%; Pred. No. 6.7e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
Db 20 ACTGATCCGCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 13
AX093047 9285 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 52 from Patent WO0118192.
ACCESSION AX093047
VERSION AX093047.1 GI:13509522
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 9285)
AUTHORS Gruber, V. and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods
for obtaining them
JOURNAL Patent: WO 0118192-A 52 15-MAR-2001;
MERISTEM THERAPEUTICS (FR)
FEATURES Location/Qualifiers
source 1. .9285
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="MPr11336"

misc_feature 1

WMRFKDEYONLAKIANQDGTQAILSQIRIFLGEDPVLGONTVONIAPRKQLQVC
PNYSIRRYLMDYLTTLAETGLMSETEGPAISBELFTKMPAIGERVAOAYKIMPT
SAVILPSRVYFTINYLTAEQCEASVMSLKAIDECRDPFIEGYGRSEKAYTAKA
TKYTKRAHDHIRTAKYORKCKCYI GOGHYANOCRNKHQDOORVALIOSLDLKE
NEBYGADDDKEEDDEIFSVLGEDYOEETIMWI EEDDIOOI I KESKRGDLSRRNG
PNFPGAEVQGVLPKPSWRRPIQATLEBNCHNMTAISTQLACRSCKQPLAGVO
CHCHAVYCFMCAEAYHDVQAEKILSKDYSPSARGKGAVII EDEI BGEPLISOLQ
OENORLOKOVERLOEBELKLRKEDBALKHSBKASRYFTIOESDEALNLIKELRO
FKEETRMALIAQLKALIVOEEDTIEERCAMILEEKHTENITYSATAKAENGLYVKG
IKDNNMEPYIINALVDGTACLIQISAI PENYEDAKVTNPRSVIGTSTOMIKA
GRILLISQYFRMPTYVMNGLSPGIOMIIGCSFIRSLBEGRLI EKDITTPYKLVTSI
ETBRTIOVANSIEBELSEDEYINLAASVETPSFLDOEFARKNDLLEKEMKTYGE
NPMEFWNNKIKCKLNI INDIKIMGRPIKGVTPGDEAMTROINLLQMKVIRPSES
KHSRTAFIVRSGETIDPITGKEKKERWENYKLI NENTESDQYSLPGINTIISKVG
RSKIYSEPLDKSGFMVAMEBESVVTAPLAGNLYEMLVMPGLKNAPAI FORKDN
VPFGTEKEIYIVYIDDIIVFSETAEOSHLYTMQLCKENGILISPTMKIGTPEIDP
LGASLGCTKIKLOPHITISKIDPSDEKLATPEGRSWLGLISTARANYIODIGKLOPL
KOKAPFTGDYKMNPEYKMWVQIKERKYNLPDLQLPKDSFII IETDGMWGAVCK
WMSKIDPRSTERICAYASGFNPDKSTIDAEIOAITHGLDKPKIYVLDKRELIRSD
CEALIKFYNNTNENKPSRVWLFPSPDLTGLGITVTEPHIDGKHNGIADALSRMINPI
VERKDESPYFTSSVEDALKVNCNDHGRNLISAVINDIITVLR.

variation

variation

old_sequence

variation

variation

variation

variation

variation

variation

variation

variation

variation

BASE COUNT 2662 a 1389 c 1577 g 1861 t
ORIGIN

Query Match 72.3%; Score 47; DB 14; Length 7489;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TATCCGCGCATCATGATCATGACAGTACTGAGAGATGAATA 62

Db 7123 TATCCGCGCATCATGATCATGACAGTACTGAGAGATGAATA 7169

Search completed: May 11, 2003, 06:04:51
Job time : 695.446 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 126.416 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	65	21	AAA96842
2	65	100.0	317	21	AAA96837
3	65	100.0	371	21	AAA96839
4	58	89.2	348	21	AAA96838
5	58	89.2	398	21	AAA96840
6	58	89.2	472	21	AAA96859
7	58	89.2	541	21	AAA96858
8	58	89.2	604	21	AAA96857
9	55	84.6	392	21	AAA96855

10	55	84.6	393	21	AAA96853	Nucleotide sequenc
11	55	84.6	462	21	AAA96854	Nucleotide sequenc
12	55	84.6	600	21	AAA96856	Nucleotide sequenc
13	46	70.8	243	21	AAA96835	Promoter from inte
14	26.4	40.6	2837	23	AB111216	Drosophila melanog
15	24.8	38.2	5668	23	AB114687	Drosophila melanog
16	24.8	38.2	9477	23	AB114686	Drosophila melanog
17	24.4	37.5	1006	21	AC33441	Arabidopsis thalia
18	24.2	37.2	2600	17	AA743624	Chromatin regulato
19	24.2	37.2	2600	14	AB165847	Lung cancer relate
20	24.2	37.2	22798	22	AA681124	Human immune/haema
21	24.2	37.2	22798	22	AA689195	Human digestive sy
22	24	36.9	452	23	AA584412	DNA encoding novel
23	24	36.9	569	23	ABV54779	Human prostate exp
24	24	36.9	755	22	AAH75159	Nucleotide sequenc
25	24	36.9	1470	22	AA158954	Human polynucleoti
26	24	36.9	2320	21	AA37735	Arabidopsis thalia
27	23.8	36.6	1467	23	AA584414	DNA encoding novel
28	23.8	36.6	30115	21	AA59651	Modified adenovira
29	23.6	36.3	273	22	AA524051	Human ovarian PCR-
30	23.6	36.3	385	23	AAH82620	Human ovarian tumo
31	23.6	36.3	1863	23	AB121739	Drosophila melanog
32	23.6	36.3	2884	23	AB121738	Drosophila melanog
33	23.6	36.3	4168	24	AB192106	Human Tumour Endot
34	23.4	36.0	424	22	AB14122	Human breast cance
35	23.4	36.0	468	22	AA12735	Human breast cance
36	23.4	36.0	570	22	AA12984	Human breast cance
37	23.4	36.0	764	22	AA121507	Human breast cance
38	23.4	36.0	147724	22	ABK83566	Human CDNA differe
39	23.2	35.7	470	22	AAH31993	Human olfactory re
40	23	35.4	1809	21	AA359514	Human secreted pro
41	22.8	35.1	480	21	AA359584	Eucalyptus grandis
42	22.8	35.1	349980	22	AA41324	Pyrococcus abyssi
43	22.6	34.8	459	22	ABH44936	Human breast cell
44	22.6	34.8	459	22	ABH55410	Human foetal liver
45	22.6	34.8	459	22	ABH25138	Probe #3604 for ge

ALIGNMENTS

RESULT 1	AAA96842	standard; DNA; 65 BP.
ID	AAA96842	
AC	AAA96842;	
XX		
DT	19-FEB-2001	(first entry)
XX		
DE	Directional desoxynucleotide building block SL.	
XX		
KW	Promoter, intergenic region; Comelina yellow mottle virus;	
KW	chimeric expression promoter; plant vascular expression promoter;	
KW	plant green tissue expression promoter; Cassava vein mosaic virus;	
KW	transgenic plant; ss.	
XX		
OS	Synthetic.	
XX		
PN	WO200058485-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	29-MAR-2000; 2000WO-IB00370.	
XX		
PR	29-MAR-1999; 99FR-0003925.	
XX		
PA	(MERI-) MERISTEM THERAPEUTICS.	
XX		
PI	Rance I, Gruber V, Theisen M;	
DR	WPI; 2000-647238/62.	
XX		
PT	Chimeric expression promoter for transgenic plant production, comprises	

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX Disclosure; Page 23; 91pp; English.
PS
CC The present sequence represents a directional desoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comellina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic plants. The chimeric promoters are useful
CC for producing transgenic plants.

SO Sequence 65 BP; 20 A; 16 C; 14 G; 15 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 65;

Best Local Similarity 100.0%; Pred. No. 5.1e-15;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
DB 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
QY 61 TAGCT 65
DB 61 TAGCT 65

RESULT 2

AAA96837 standard; DNA; 317 BP.

QY AAA96837;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1116.

XX Promoter; intergenic region; Comellina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Comellina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

XX -05-OCT-2000.

PF 29-MAR-2000; 2000WO-IB00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PS Claim 5; Page 81; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comellina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic plants. The chimeric promoters are useful
CC for producing transgenic plants.

SO Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 317;

Best Local Similarity 100.0%; Pred. No. 8e-15; Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
DB 8 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 67
QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 3

AAA96839 standard; DNA; 371 BP.

QY AAA96839;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1116.

XX Promoter; intergenic region; Comellina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Comellina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

XX -05-OCT-2000.

PF 29-MAR-2000; 2000WO-IB00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

PT Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 81; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 8.3e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTGACAGTATGATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAA 60
DB 8 CATCTGACAGTATGATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAA 67

QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 4

AAA96838 standard; DNA; 348 BP.

AC AAA96838;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr1117.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

PI WPI; 2000-647238/62.

XX Claim 5; Page 81; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

Query Match 89.2%; Score 58; DB 21; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 65
DB 49 CAGACTAGTATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 106

RESULT 5

AAA96840 standard; DNA; 398 BP.

AC AAA96840;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr1147.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

PI WPI; 2000-647238/62.

XX Claim 5; Page 82; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

XX for producing transgenic plants.

XX Sequence 398 BP; 128 A; 80 C; 93 G; 97 T; 0 other;

Query Match 89.2%; Score 58; DB 21; Length 398;
Best Local Similarity 100.0%; Pred. No. 3.7e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 65
DB 49 CAGACTAGTATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 106

RESULT 6

AAA96859

ID	AA96859 standard; DNA; 472 BP.
XX	
XX	AA96859;
AC	
XX	
DT	19-FEB-2001 (first entry)
XX	
XX	
DE	Nucleotide sequence of chimeric expression promoter MP1169.
XX	
KM	Promoter; intergenic region; Commelina yellow mottle virus;
KM	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; chimera; ss.
XX	
OS	Chimeric - Commelina yellow mottle virus.
OS	Chimeric - Cassava vein mosaic virus.
XX	
PN	MO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
PP	29-MAR-2000; 2000MO-1B00370.
XX	
PR	29-MAR-1999; 99PR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
P1	Rance I, Gruber V, Thiesen M;
DR	WPI; 2000-647238/62.
XX	
PT	Chimeric expression promoter for transgenic plant production, comprises
PT	sequence from promoter comprising vascular expression region replaced
PT	with sequence from promoter comprising green tissue expression region
XX	
PS	Claim 5; Page 88; 91pp; English.
XX	
CC	The present sequence represents a chimeric promoter of the invention.
CC	The specification describes chimeric expression promoters. These
CC	chimeric promoters comprise a nucleic acid sequence which is derived
CC	from a first plant promoter, in which a plant vascular expression
CC	promoter region is replaced with a nucleic acid sequence derived from
CC	a second plant promoter comprising a plant green tissue expression
CC	promoter region. Preferably, the first plant promoter originates from
CC	Commelina yellow mottle virus, and the second plant promoter originates
CC	from the Cassava vein mosaic virus. Especially, the promoters are
CC	derived from intergenic regions. The chimeric promoters are useful
CC	for producing transgenic plants.
XX	
SO	Sequence 472 BP, 149 A, 92 C, 112 G, 119 T, 0 other;
	Query Match 89.2%; Score 58; DB 21; Length 472;
	Best Local Similarity 100.0%; Pred. No. 3.9e-12;
	Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	8 CAGACTAGTATCCGCCGTCATCATGACATCATCACAGTACTGAGAGATGAATAGCT 65
DB	49 CAGACTAGTATCCGCCGTCATCATGACATCATCACAGTACTGAGAGATGAATAGCT 106
RESULT 7	
ID	AAA96858
XX	AAA96858 standard; DNA; 541 BP.
AC	
XX	AAA96858;
XX	
DT	19-FEB-2001 (first entry)
XX	
XX	
DE	Nucleotide sequence of chimeric expression promoter MP1168.
XX	
KM	Promoter; intergenic region; Commelina yellow mottle virus;
KM	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;

KW	transgenic plant; chimera; ss.
OS	Chimeric - Commelina yellow mottle virus.
OS	Chimeric - Cassava vein mosaic virus.
PN	MO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
XX	29-MAR-2000; 2000MO-IB00370.
XX	
PR	29-MAR-1999; 99PR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
P1	Rance I, Gruber V, Theisen M;
XX	
DR	WPI; 2000-647238/62.
XX	
P7	Chimeric expression promoter for transgenic plant production, comprises
P7	sequence from promoter comprising vascular expression region replaced
P7	with sequence from promoter comprising green tissue expression region
XX	
PS	Claim 5; Page 87-88; 91pp; English.
CC	The present sequence represents a chimeric promoter of the invention.
CC	The specification describes chimeric expression promoters. These
CC	chimeric promoters comprise a nucleic acid sequence which is derived
CC	from a first plant promoter, in which a plant vascular expression
CC	promoter region is replaced with a nucleic acid sequence derived from
CC	a second plant promoter comprising a plant green tissue expression
CC	promoter region. Preferably, the first plant promoter originates from
CC	Commelina yellow mottle virus, and the second plant promoter originates
CC	from the Cassava vein mosaic virus. Especially, the promoters are
CC	derived from intergenic regions. The chimeric promoters are useful
CC	for producing transgenic plants.
XX	
SQ	Sequence 541 BP, 169 A; 104 C; 130 G; 138 T; 0 other;
XX	
Query Match	89.2%; Score 58; DB 21; Length 541;
Best Local Similarity	100.0%; Pred. No. 4e-12;
Matches	58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAACAGTACTGAGAGATGAATGCT 65
DB	49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAACAGTACTGAGAGATGAATGCT 106
RESULT 8	
AAA96857	
ID	AAA96857 standard; DNA; 604 BP.
XX	
AC	AAA96857;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Nucleotide sequence of chimeric expression promoter MP1167.
XX	
KW	Promoter; intergenic region; Commelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; chimera; ss.
XX	
OS	Chimeric - Commelina yellow mottle virus.
OS	Chimeric - Cassava vein mosaic virus.
XX	
PN	MO200058485-A1.
XX	
XX	05-OCT-2000.
PD	
XX	
PF	29-MAR-2000; 2000MO-IB00370.
XX	

PR 29-MAR-1999; 99FR-0003925.
XX (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
PS Claim 5; Page 87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
XX
Query Match 89.2%; Score 58; DB 21; Length 604;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 8 CAGCTGATATCCCGCTCATCATGATCATCATCATCTAGAGATGAATGCT 65
DB 49 CAGACTGATATCCCGCTCATCATGATCATCATCATCTAGAGATGAATGCT 106
XX
RESULT 9
AAA96855
ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1164.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT

PT -
XX
PS Claim 5; Page 86; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 84.6%; Score 55; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 11 ACTGATATCCCGCTCATCATGATCATCATCATGATGAGATGAATGCT 65
DB 20 ACTGATATCCCGCTCATCATGATCATCATCATGATGAGATGAATGCT 74
XX
RESULT 10
AAA96853
ID AAA96853 standard; DNA; 393 BP.
XX
AC AAA96853;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1162.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
PS Claim 5; Page 85; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;

Query Match 84.6%; Score 55; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ACTGATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
DB 20 ACTGATATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 11

AAA96854
ID AAA96854 standard; DNA; 462 BP.

XX AAA96854;

XX 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr1163.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

KM transgenic plant; chimera; 88.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX WO200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

XX PT sequence from promoter comprising vascular expression region replaced

XX PT with sequence from promoter comprising green tissue expression region

XX Claim 5; Page 86; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

XX for producing transgenic plants.

XX Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;

Query Match 84.6%; Score 55; DB 21; Length 462;

Best Local Similarity 100.0%; Pred. No. 5.2e-11;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ACTGATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
DB 20 ACTGATATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 12

AAA96856
ID AAA96856 standard; DNA; 600 BP.

XX AAA96856;

XX 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr165.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

KM transgenic plant; chimera; 88.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX WO200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

XX PT sequence from promoter comprising vascular expression region replaced

XX PT with sequence from promoter comprising green tissue expression region

XX Claim 5; Page 86-87; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

XX for producing transgenic plants.

XX Sequence 600 BP; 188 A; 111 C; 147 G; 154 T; 0 other;

Query Match 84.6%; Score 55; DB 21; Length 600;

Best Local Similarity 100.0%; Pred. No. 5.6e-11;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX PS Claim 1; SEQ ID NO 38543; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 5668 BP; 1478 A; 1411 C; 1420 G; 1359 T; 0 other;

Query Match 38.2%; Score 24.8; DB 23; Length 5668;

Best Local Similarity 63.3%; Pred. No. 26;

Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 CATGCTGCAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAA 60
DB 667 CATCTCTTCATGATGATGCGCTTCATCTTCGACTTCATCTCAATGTGAAATCTGCA 608

Search completed: May 11, 2003, 03:04:00
Job time : 128.416 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 26.0652 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65

Sequence: 1 catgctgcagactatcc.....tactgagagatgatcatc 65

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22.2	34.2	1086	US-09-183-861-77	Sequence 77, Appl
C 2	22.2	34.2	1086	US-09-022-765-77	Sequence 77, Appl
C 3	22.2	34.2	1179	US-09-147-926-1	Sequence 1, Appl
C 4	22.2	33.8	4010	US-08-785-310A-3	Sequence 3, Appl
C 5	21.8	33.5	1515	US-09-134-001C-143	Sequence 143, App
C 6	21.8	33.5	5393	US-08-591-079-9	Sequence 9, Appl
C 7	21.8	33.5	13473	PCT-US96-03916-1	Sequence 1, Appl
C 8	21.8	33.5	18912	PCT-US96-03916-59	Sequence 59, Appl
C 9	21.6	33.2	4940	US-08-484-105-1	Sequence 1, Appl
C 10	21.6	33.2	4940	US-08-484-106-1	Sequence 1, Appl
C 11	21.6	33.2	8561	US-09-112-450-3	Sequence 3, Appl
C 12	21.6	33.2	8561	US-09-419-281A-3	Sequence 3, Appl
C 13	21.6	32.6	1001	US-09-641-638-111	Sequence 111, App
C 14	21.2	32.6	1001	US-09-641-638-112	Sequence 112, Appl
C 15	21.2	32.6	3116	US-09-362-831-10	Sequence 10, Appl
C 16	21.2	32.6	7082	US-09-362-831-1	Sequence 1, Appl
C 17	21.2	32.6	1525	US-08-186-833-3	Sequence 3, Appl
C 18	21.2	32.3	2085	US-09-289-843A-65	Sequence 65, Appl
C 19	21.2	32.3	2085	US-09-088-357B-65	Sequence 65, Appl
C 20	21.2	32.3	4403765	US-09-103-840A-2	Sequence 2, Appl
C 21	21.2	32.3	4411529	US-09-103-840A-1	Sequence 1, Appl
C 22	20.8	32.0	1010	US-09-453-702B-4	Sequence 4, Appl
C 23	20.8	32.0	2818	US-08-366-276-1	Sequence 1, Appl
C 24	20.8	32.0	3282	US-08-072-574-11	Sequence 11, Appl
C 25	20.8	32.0	3282	US-08-486-270-11	Sequence 11, Appl
C 26	20.8	32.0	3282	US-08-367-264-11	Sequence 11, Appl
C 27	20.8	32.0	3282	US-09-153-757-11	Sequence 11, Appl

C 28	20.8	32.0	4085	US-08-072-574-7	Sequence 7, Appl
C 29	20.8	32.0	4085	US-08-486-270-7	Sequence 7, Appl
C 30	20.8	32.0	4085	US-08-367-264-7	Sequence 7, Appl
C 31	20.8	32.0	4085	US-09-153-757-7	Sequence 7, Appl
C 32	20.8	32.0	4181	US-08-072-574-9	Sequence 9, Appl
C 33	20.8	32.0	4181	US-08-486-270-9	Sequence 9, Appl
C 34	20.8	32.0	4181	US-08-367-264-9	Sequence 9, Appl
C 35	20.8	32.0	4181	US-09-153-757-9	Sequence 9, Appl
C 36	20.8	32.0	4207	US-08-660-148-1	Sequence 1, Appl
C 37	20.8	32.0	4207	US-08-660-148-3	Sequence 3, Appl
C 38	20.8	32.0	4303	US-08-660-148-4	Sequence 4, Appl
C 39	20.8	32.0	4303	US-08-660-148-6	Sequence 6, Appl
C 40	20.8	32.0	7096	US-09-221-017B-373	Sequence 373, App
C 41	20.8	32.0	8257	US-09-484-970B-65	Sequence 65, Appl
C 42	20.8	32.0	46819	US-09-453-702B-72	Sequence 72, Appl
C 43	20.6	31.7	1080	US-09-147-926-3	Sequence 3, Appl
C 44	20.6	31.7	9179	US-09-453-702B-162	Sequence 162, App
C 45	20.4	31.4	463	US-08-943-731-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1
US-09-183-861-77/c
Sequence 77, Application US/09183861
Patent No. 6365165
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,861
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/022,765
FILING DATE: 12-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-183-861-77

Query Match 34.2% Score 22.2; DB 4; Length 1086;
Best Local Similarity 59.0% Pred. No. 22;
Matches 36; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Best Local Similarity 63.0%; Pred. No. 35;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCTACATGAGAG 55
DB 1829 ATCTGAGAGCATATTCGGTGGCAAGAGAGCTCCCAAGATCCAGAG 1882

RESULT 5

US-09-134-001C-143
Sequence 143, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 143
LENGTH: 1515
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-143

Query Match 33.5%; Score 21.8; DB 4; Length 1515;
Best Local Similarity 70.7%; Pred. No. 33;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCTACATCA 42
DB 497 ATGCTGAGATTCATATACGCATCATCAATGTAAATTA 537

RESULT 6

US-08-591-079-9/c
Sequence 9, Application US/08591079
Patent No. 5972899
GENERAL INFORMATION:
APPLICANT: Zychlinsky, Arturo
TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,079
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Livanac, Shmuel
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: 15661-20017.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0764
TELEX: 90-4030 MRSNPOERSWSH
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 5393 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Salmonella typhi
STRAIN: Ty2
FEATURE:

NAME/KEY: CDS
LOCATION: 543..2324
OTHER INFORMATION: /gene= "sipB"
US-08-591-079-9

Query Match 33.5%; Score 21.8; DB 2; Length 5393;
Best Local Similarity 61.4%; Pred. No. 45;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCACAGTACTGAGAGAT 57
DB 3822 CTGCTGATTTTATAGTACCGCCCTCGCATATGAAATGATATCTGTCCCGCATGTT 3766

RESULT 7

PCT-US96-03916-1/c
Sequence 1, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 13473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1059..2489
NAME/KEY: CDS
LOCATION: 2575..4107

FEATURE:
NAME/KEY: CDS
LOCATION: 4113..4445
FEATURE:
NAME/KEY: CDS
LOCATION: 4609..5487
FEATURE:
NAME/KEY: CDS
LOCATION: 5697..8654
FEATURE:
NAME/KEY: CDS
LOCATION: 9874..10962
FEATURE:
NAME/KEY: CDS
LOCATION: 11159..12658
FEATURE:
NAME/KEY: CDS
LOCATION: 12665..13447
PCT-US96-03916-1

Query Match 33.5%; Score 21.8; DB 5; Length 13473;
Best Local Similarity 65.3%; Pred. No. 56;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 16 TATCCGCGTCATCATGACATCATCACAGTACTGAGGAGATGATAGC 64
Db 11397 TTTCCTCGTGTAAATTAATCATGAAAGCAATAGTTAGCGAAGAC 11349

RESULT 8
PCT-US96-03916-59/c
Sequence 59, Application PC/TUS9603916

GENERAL INFORMATION:

APPLICANT: Wild, Martha A.

APPLICANT: Cochran, Mark D.

TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patgenin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/03916

FILING DATE: 23-MAR-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/126,597

FILING DATE: 24-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 39116-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 18912 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: linear

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

FEATURE:
NAME/KEY: CDS
LOCATION: 697..1533
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1900..2784)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2916..3605)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 3694..5124
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 5210..7081
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 7245..8123
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 8333..11290
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 11098..12402
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 12510..13598
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 13792..15291
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 15298..16080
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 16129..17013
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (17380..18216)
OTHER INFORMATION:
PCT-US96-03916-59

Query Match 33.5%; Score 21.8; DB 5; Length 18912;
Best Local Similarity 65.3%; Pred. No. 61;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 16 TATCCGCGTCATCATGACATCATCACAGTACTGAGGAGATGATAGC 64
Db 14030 TTTCCTCGTGTAAATTAATCATGAAAGCAATAGTTAGCGAAGAC 13982

RESULT 9
US-08-484-105-1/c

Sequence 1, Application US/08484105

Patent No. 5589341

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce

APPLICANT: BELL, Stephen P.

APPLICANT: KOBAYASHI, Ryuji

APPLICANT: RINE, Jasper

APPLICANT: FOSS, Margit

APPLICANT: Abad et al.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: P3393D1
CURRENT APPLICATION NUMBER: US/09/419,291A
CURRENT FILING DATE: 1999-10-15
PRIORITY FILING DATE: 1998-07-09
PRIORITY APPLICATION NUMBER: US 60/074,308
PRIORITY FILING DATE: 1998-02-11
PRIORITY APPLICATION NUMBER: US 60/052,273
PRIORITY FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 8561
TYPE: DNA
ORGANISM: Candida albicans
US-09-419-291A-3

Query Match 33.2%; Score 21.6; DB 4; Length 8561;
Best Local Similarity 85.7%; Pred. No. 60;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 25 TCATCATGACATCATCATGACTGAG 52
DB 5413 TAATCATGACACATGACATGAG 5386

RESULT 13
US-09-641-638-111/c
Sequence 111, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET 051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIORITY APPLICATION NUMBER: US 09/502,330
PRIORITY FILING DATE: 2000-02-11
PRIORITY APPLICATION NUMBER: US 60/133,200
PRIORITY FILING DATE: 1999-05-07
PRIORITY APPLICATION NUMBER: US 09/275,267
PRIORITY FILING DATE: 1999-03-23
PRIORITY APPLICATION NUMBER: US 60/119,917
PRIORITY FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 111
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-406-52 : polymorphic base C or T
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-406-52.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-406-52.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 450..468
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 881..889
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding

LOCATION: 489..513
OTHER INFORMATION: 12-406-52 potential probe
NAME/KEY: misc feature
LOCATION: 4,53,104,243,324,369
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-111

Query Match 32.6%; Score 21.2; DB 4; Length 1001;
Best Local Similarity 69.0%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24 GTCAATGACATCATCATGAGATGAGATGACT 65
DB 845 GTGAAAATTAACATCAAAACAGAGCTGAGAGGATTTCT 804

RESULT 14
US-09-641-638-112/c
Sequence 112, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET 051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIORITY APPLICATION NUMBER: US 09/502,330
PRIORITY FILING DATE: 2000-02-11
PRIORITY APPLICATION NUMBER: US 60/133,200
PRIORITY FILING DATE: 1999-05-07
PRIORITY APPLICATION NUMBER: US 09/275,267
PRIORITY FILING DATE: 1999-03-23
PRIORITY APPLICATION NUMBER: US 60/119,917
PRIORITY FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 112
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-406-409 : polymorphic base A or G
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-406-409.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-406-409.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 93..111
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 524..542
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 489..513
OTHER INFORMATION: 12-406-409 potential probe
NAME/KEY: misc feature
LOCATION: 12,717,742,755,758,1000..1001
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-112

Query Match 32.6%; Score 21.2; DB 4; Length 1001;
Best Local Similarity 69.0%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24 GTCAATGACATCATCATGAGATGAGATGACT 65

Db 488 GTGAAATATACATCAAAACAGAGCTGAGAGGAAATTCT 447

RESULT 15

US-09-362-831-10/c
; Sequence 10, Application US/09362831
; Patent No. 6306400
; GENERAL INFORMATION:
; APPLICANT: BUBLOT et al.
; TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING, AS VECTOR, THE
; FILE REFERENCE: 454313-2520
; CURRENT APPLICATION NUMBER: US/09/362,831
; CURRENT FILING DATE: 1999-07-28
; NUMBER OF SEQ. ID NOS: 28
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3116
; TYPE: DNA
; ORGANISM: Infectious Laryngotracheitis Virus
US-09-362-831-10

Query Match 32.6%; Score 21.2; DB 4; Length 3116;
Best Local Similarity 64.0%; Pred. No. 66;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTATATCCGCGCATCAATGACATCATCAGTACTGA 51
Db 2153 ATGCTGAGAGAGATTCCTTATCATCATGATATCATCAATCATCA 2104

Search completed: May 11, 2003, 03:07:06
Job time : 38.0652 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1044.89 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65

Sequence: 1 catgctgcagactagatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_estum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28.4	43.7	614	12	BF168473 601776065
C 2	27.4	42.2	563	17	BH372871 AG-ND-106
C 3	27.4	42.2	665	17	BH365471 AG-ND-106
C 4	27.4	42.2	814	17	BH371422 AG-ND-172
C 5	27	41.5	785	13	BH971380 GM830013A
C 6	26.8	41.2	769	12	BG433137 602496882

C 7	26.8	41.2	968	11	BC034140	BC034140 Homo sapi
C 8	26.4	40.6	543	12	BE975786	BE975786 b945c02.x
C 9	26.2	40.3	772	17	CNS015AK	AL164981 Tetradodon
C 10	26	40.0	103	12	BF466720	BF466720 UI-M-CGDP
C 11	26	40.0	135	14	BM942017	BM942017 UI-M-CGDP
C 12	26	40.0	482	10	BE531686	BE531686 601230802
C 13	26	40.0	545	14	BQ562476	BQ562476 H4076B12-
C 14	26	40.0	549	13	B1713338	B1713328 IC85H09.Y
C 15	26	40.0	566	13	B1715981	B1715981 IC64D03.Y
C 16	26	40.0	613	10	BB621994	BB621994 BB621994
C 17	26	40.0	617	10	BB659983	BB659983 BB659983
C 18	26	40.0	618	10	BB618417	BB618417 BB618417
C 19	26	40.0	618	10	BB651874	BB651874 BB651874
C 20	26	40.0	642	10	BB655827	BB655827 BB655827
C 21	26	40.0	654	10	BB638327	BB638327 BB638327
C 22	26	40.0	663	12	BF348858	BF348858 RCL-DT002
C 23	26	40.0	703	9	AI1747024	AI1747024 U112604.Y
C 24	26	40.0	748	13	BG969043	BG969043 602834988
C 25	26	40.0	768	14	BQ745618	BQ745618 UI-M-EMO-
C 26	26	40.0	868	13	B1739607	B1739607 603361862
C 27	26	40.0	938	12	BG293515	BG293515 602390444
C 28	26	40.0	954	14	BQ884980	BQ884980 AGENCOURT
C 29	26	40.0	973	14	BQ898008	BQ898008 AGENCOURT
C 30	25.8	39.7	388	9	AI344310	AI344310 TC03d11.x
C 31	25.8	39.7	388	9	AI344331	AI344331 TC03f12.x
C 32	25.8	39.7	656	17	AG064170	AG064170 Pan trogl
C 33	25.6	39.4	515	13	BM052014	BM052014 Tc ad2.02
C 34	25.6	39.4	547	14	BQ097606	BQ097606 ph05a07.Y
C 35	25.6	39.4	930	17	CNS07CK4	AL439742 T7 end of
C 36	25.4	39.1	308	17	AZ218097	AZ218097 Sheared D
C 37	25.4	39.1	524	17	AZ217624	AZ217624 Sheared D
C 38	25.2	38.8	767	12	BE748221	BE748221 601571441
C 39	25	38.5	379	9	AA235966	AA235966 z805f02.s
C 40	25	38.5	381	9	AA814930	AA814930 oc07c03.s
C 41	25	38.5	438	10	AW971534	AW971534 B5T383623
C 42	24.8	38.2	495	14	BQ378960	BQ378960 RC3-UT006
C 43	24.8	38.2	504	9	AA542644	AA542644 fa08e10.r
C 44	24.8	38.2	530	9	AA390522	AA390522 LD08708.5
C 45	24.8	38.2	619	17	AZ952780	AZ952780 2M0217C18

ALIGNMENTS

RESULT 1
LOCUS BF168473/c 614 bp mRNA linear EST 30-OCT-2000
DEFINITION 601776065F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4017668 5',
mRNA sequence.
ACCESSION BF168473
VERSION BF168473.1 GI:11048825
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 614)
NHI-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strauberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM9267 row: a column: 17
High quality sequence stop: 614.
Location/Qualifiers 1. 614

FEATURES
source

ORIGIN

Query Match 43.7% Score 28.4; DB 12; Length 614;
 Best Local Similarity 65.1% Pred. No. 18;
 Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

DB 173 TTCTGCACTAGTATCCGCGCTCATCATGATCATCATGATGAGATGATGATA 62
 /clone_1lb="NCI_CGAP_Lu29"
 /clone_lib="NCI_CGAP_Lu29"
 /lab host="DH108"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 198 a 97 c 154 g 164 t 1 others

Query Match 43.7% Score 28.4; DB 12; Length 614;
 Best Local Similarity 65.1% Pred. No. 18;
 Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

DB 173 TTCTGCACTAGTATCCGCGCTCATCATGATCATCATGATGAGATGATGATA 62
 /clone_1lb="NCI_CGAP_Lu29"
 /clone_lib="NCI_CGAP_Lu29"
 /lab host="DH108"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 198 a 97 c 154 g 164 t 1 others

RESULT 2
 BH372871/c 563 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-106G18.TR ND-TAM Anopheles gambiae genomic clone AG-ND-106G18
 DEFINITION , DNA sequence.
 ACCESSION BH372871 GI:173319013
 VERSION BH372871
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anophelinae.

REFERENCE 1 (bases 1 to 563)
 AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AG-ND-106G18.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

FEATURES

source

1..563
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-106G18"
 /clone_1lb="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 157 a 107 c 83 g 216 t

ORIGIN

Query Match 42.2% Score 27.4; DB 17; Length 563;
 Best Local Similarity 65.6% Pred. No. 38;
 Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

DB 342 ATTTGCAAGAGCTTAAGACCTTACGAGAGCAATATGATTAAGAGAAATATAT 283
 /clone_1lb="ND-TAM"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 196 a 119 c 100 g 250 t

RESULT 3
 BH385471/c 665 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-106B1.TR ND-TAM Anopheles gambiae genomic clone AG-ND-106B1.
 DEFINITION , DNA sequence.
 ACCESSION BH385471 GI:17331613
 VERSION BH385471
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anophelinae.

REFERENCE 1 (bases 1 to 665)
 AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AG-ND-106B1.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

FEATURES

source

1..665
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-106B1"
 /clone_1lb="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 196 a 119 c 100 g 250 t

Query Match 42.2% Score 27.4; DB 17; Length 665;
 Best Local Similarity 65.6% Pred. No. 41;
 Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

DB 341 ATTTGCAAGAGCTTAAGACCTTACGAGAGCAATATGATTAAGAGAAATATAT 282
 /clone_1lb="ND-TAM"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 196 a 119 c 100 g 250 t

RESULT 4	BH371422/c	BH371422	814 bp	DNA	linear	GSS 10-DEC-2001
LOCUS	AG-ND-172C15.TF	ND-TAM	Anopheles gambiae	genomic clone	AG-ND-172C15	
DEFINITION	DNA sequence.					
ACCESSION	BH371422					
VERSION	BH371422.1	GI:17317547				
KEYWORDS	GSS.					
SOURCE	African malaria mosquito.					
ORGANISM	Anopheles gambiae					
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.					
AUTHORS	1 (bases 1 to 814)					
TITLE	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.					
JOURNAL	Direct Submission of BAC-end sequences from Anopheles gambiae unpublished (2001)					
COMMENT	Other GSSs: AG-ND-172C15.TF					
	Contact: Brendan J Loftus					
	Department of Eukaryotic Genomics					
	The Institute for Genomic Research					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel: 301 838 0208					
	Fax: 301 838 3543					
	Email: b.loftus@tigr.org					
	This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.					
	Seq primer: M13 For					
	Class: BAC ends.					
FEATURES	Location/Qualifiers					
source	1..814					
	/organism="Anopheles gambiae"					
	/strain="PEST"					
	/db_xref="taxon:7165"					
	/clone="AG-ND-172C15"					
	/clone_11b="ND-TAM"					
	/note="vector: pECBAC1; Site_1: HindIII"					
BASE COUNT	237 a	141 c	132 g	304 t		
ORIGIN						
Query Match	42.2%	Score 27.4;	DB 17;	Length 814;		
Best Local Similarity	65.6%	Pred. No. 45;				
Matches 40;	Conservative 0;	Mismatches 21;	Indels 0;	Gaps 0;		
CY	2	ATGCGACACACTAGTATCGCGCGTATCATATGACATCATCAGTACTAGAGAGATGAT	61			
DB	347	ATTTCGACGACAGATTAAGACCTTACGGAAGACATATTCAGATTACAGAGAAATGAT	288			
CY	62	A	62			
DB	287	A	287			
RESULT 5	B1971380	785 bp	mRNA	linear	EST 23-OCT-2001	
LOCUS	GMS30013A20F02	Gm-r1083	Glycine max	CDNA clone	Gm-r1083-4852 3'	
DEFINITION	mRNA sequence.					
ACCESSION	B1971380					
VERSION	B1971380.1	GI:16345785				
KEYWORDS	EST.					
SOURCE	soybean.					
ORGANISM	Glycine max					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Spermatophytes: Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 785)

Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Corryell,V., Epling,J., Raph,C., Shoop,E., Padias,J., Liu,L. and Lewin,H.

A Functional Genomics Program for Soybean (NSF 9872565)

Unpublished (1999)

Other ESTs: BE022118 corresponding to Gm-cl028-9003 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Incyte Genomics, 4633 World Parkway Circle St. Louis, Missouri 63134. Phone (800) 430-0030 or (314) 427-3222 FAX: (314) 427-3324. Web site: <http://www.incyte.com/reagents/catalog.jsp?page=clones/collaboratio>
n/index

Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES
Source

Location/Qualifiers

1..785
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1083-4852"
/clone_id="Gm-r1083"
/note="The library Gm-r1083 is a sequence-driven, rerecked set of 4,992 clones selected from cDNA libraries from various tissues and stages of development of soybean. It represents 1117 sequences from the progenitor library Gm-cl1009 (from mature roots of 2 month old greenhouse grown 'Williams' soybean plants); 820 sequences from the progenitor library Gm-cl013 (from 2 to 3 week old whole plants of Williams) and 3055 sequences from library Gm-cl028 (from 'Superpod' plants whose seedlings were inoculated with *Bradyrhizobium japonicum*, courtesy of Dr. Gary Stacey). The 5' ESTs of the source clones from the different progenitor libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1083. The cDNA clones of the rerecked Gm-r1083 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://web.ahc.umn.edu/biodata/nefsoy/>. Rerecking was performed by Incyte Genomics, St. Louis, <http://www.incyte.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.lie.uiuc.edu/bioelec/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1083 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Incyte Genomics clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 220 a 177 c 134 g 247 t 7 others

ORIGIN

Query Match 41.5%; Score 27; DB 13; Length 785;
Best Local Similarity 66.1%; Pred. No. 61;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

3 TGCTGCAAGTATGATCCGCGTATCATGATCATCATCAAGTATGAGATGAT 61
|||||
Db 269 TGCTTACGCTTCATCTTACCGCGCATGATGATCATTAAGTACTCTGAGCTTCA 327
|||||

LOCUS	BCg33137	769 bp	mRNA	linear	EST 14-MAR-2000
DEFINITION	602496882p1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4610636 5', mRNA sequence.				
ACCESSION	BCg33137				
VERSION	BCg33137.1	GI:13339643			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia, Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (Baeas 1 to 769)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgsab@remail.nih.gov				
	Tissue Procurement: CLONTECH Laboratories, Inc.				
	cDNA Library Preparation: CLONTECH Laboratories, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov				
	Plate: LUCM1356 row: c column: 21				
	High quality sequence stop: 741.				
FEATURES	Location/Qualifiers				
source	1..769				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4610636"				
	/clone_lib="NIH_MGC_75"				
	/lab_host="DH10B (T1 phage-resistant)"				
	/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgccgcgcgcgc); Site 2: SfiI (ggccatctagcc); 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCAGAGCCCGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC library."				
BASE COUNT	215 a 122 c 178 g 253 t 1 others				
ORIGIN					
Query Match	41.2%; Score 26.8; DB 12; Length 769;				
Best Local Similarity	64.5%; Pred. No. 70;				
Matches 40; Conservative	0; Mismatches 22; Indels 0; Gaps 0;				
QY	1	CATGTCGACGACTAATATCCGCGTCATCAATAGCATCTATCAGTACGAGAGATGA	60		
DB	390	CATTACACAGAACGATATGACACACAGGACCTGTATAAGTATGATTAATGA	331		
QY	61	TA 62			
DB	330	TA 329			
RESULT 7					
LOCUS	BC034140	968 bp	mRNA	linear	HTC 08-JUL-2002
DEFINITION	Homo sapiens, Similar to likely ortholog of yeast ARV1, clone IMAGE:4610636, mRNA.				
ACCESSION	BC034140				
VERSION	BC034140.1	GI:21707881			
KEYWORDS	HTC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia, Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (Baeas 1 to 968)				
TITLE	Strausberg, R.				
	Direct Submission				

JOURNAL

Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

REMARK

Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES

Source

1. 968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4610636"
/tissue_type="Kidney"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B"
/note="Vector: PDNR-LIB"

BASE COUNT 358 a 135 c 196 g 279 t

ORIGIN

Query Match 41.2% Score 26.8; DB 11; Length 968;
Best Local Similarity 64.5%; Pred. No. 78;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 CATGTCGACAGTAGTATCCGCGCTCATCATCATCATCATCATCTGAGAGATCAA 60
Db 389 CATTAACAGAAAGTATCATGACACACAGCAGGACCTTGATTAAGTAAATCA 330

Oy 61 TA 62
Db 329 TA 328

RESULT 8
BS975786/c 543 bp mRNA linear EST 04-OCT-2000

LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BS975786 b945c02.x1 Drosophila melanogaster adult testis library Drosophila melanogaster cDNA clone b945c02.3', mRNA sequence.
BS975786
BS975786.1 GI:10604624
EST.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 543)
Andrews, J., Bouffard, G. and Oliver, B.
Drosophila melanogaster testis expressed sequence tags
Unpublished. (1999)
Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NINDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliver@helix.nih.gov,
<http://www.nindk.nih.gov/intram/people/boliver.htm>
Tissue isolation and library construction performed at the National

TITLE Luo, A.G. and Ko, M.S.H.
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
JOURNAL Unpublished (2002)
COMMENT Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: chna@igun.gic.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://igun.gic.nia.nih.gov/cDNA/NIA_7_4k.html for details.
Plate: H4076 Row: B Column: 12
Seq primer: -21M13 Reverse
High quality sequence stop: 545
POLYA=No.

FEATURES

source Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="nia:EST:H4076B12-5"
/db_xref="taxon:10090"
/clone="H4076B12"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 176 a 113 c 104 g 152 t
ORIGIN

Query Match 40.0%; Score 26; DB 14; Length 545;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 4 GCTGCAGACTAGTATCCGCGTCATCATGATCATCATGACTGAG 53
Db 160 GCTGCAGATATGGAAGCCGCGCATGATGATGCTCACCAGACTATGTGG 111

RESULT 14 549 bp mRNA linear EST 12-MAR-2002
LOCUS B1713238/c
DEFINITION Musculus cDNA clone IMAGE:5660776 5' similar to TR:Q9VZ81 Q9VZ81
CG13708 PROTEIN.; mRNA sequence.
ACCESSION B1713238 GI:15688933
VERSION B1713238
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 549)
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, V., Williams, T.,
Jackson, Y., and Bowers, Y.
Other ESTs: ic65h09.xl
Unpublished (2000)
TITLE Endocrine Pancreas Consortium
JOURNAL
COMMENT Other ESTs: ic65h09.xl
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@iobp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu)
MGI:1947102 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 429.
Location/Qualifiers

FEATURES

source 1..549
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5660776"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMSI"
/sex="Both for embryonic & newborn, male for adult and
adult islet."
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."
BASE COUNT 163 a 131 c 110 g 145 t
ORIGIN

Query Match 40.0%; Score 26; DB 13; Length 549;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 4 GCTGCAGACTAGTATCCGCGTCATCATGATCATCATGACTGAG 53
Db 256 GCTGCAGATATGGAAGCCGCGCATGATGATGCTCACCAGACTATGTGG 207

RESULT 15 566 bp mRNA linear EST 12-MAR-2002
LOCUS B1715981/c
DEFINITION Musculus cDNA clone IMAGE:5658197 5' similar to TR:Q9VZ81 Q9VZ81
CG13708 PROTEIN.; mRNA sequence.
ACCESSION B1715981 GI:15691676
VERSION B1715981
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 566)
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, V., Williams, T.,
Jackson, Y., and Bowers, Y.
Other ESTs: ic65h09.xl
Unpublished (2000)
TITLE Endocrine Pancreas Consortium
JOURNAL
COMMENT Other ESTs: ic65h09.xl
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)

MGI:1944523 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov

Seq primer: -408P from Gibco

High quality sequence atop: 428.

Location/Qualifiers

1..566

/organism="Mus musculus"

/strain="ICR"

/db_xref="taxon:10090"

/clone="IMAGE:5658197"

/clone_lib="Melton Normalized Mixed Mouse Pancreas 1

N1-MMS1"

/sex="both for embryonic & newborn, male for adult and

adult islet"

/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,

adult, mixed"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five

libraries representing E10.5/12.5 pancreatic bud, E16.5

pancreas, newborn pancreas, adult pancreas, and adult

islets of Langerhans were separately constructed using

SuperScript Plasmid Library kit (Life Technologies). cDNA

was made by oligo-dT priming and size-selected by column

fractionation. Libraries were amplified once on solid

support and plasmid DNA from each library was prepared

and mixed in equal amounts. The mixed library DNA was

normalized by method #4 from Bonaldo, Lennon, and Soares

1996 Genome Research 6:791-806; 0.5 microgram

single-stranded mixed library plasmid DNA was mixed with

5 micrograms PCR product representing mixed library

inserts and hybridized to an EcoT of 6. Single-stranded

(unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

BASE COUNT 180 a 121 c 113 g 152 t

ORIGIN

Query Match 40.0%; Score 26; DB 13; Length 566;

Best Local Similarity 70.0%; Pred. No. 1.2e+02;

Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 GCTGAGACTAGTATCCCGCTCATCATCATCATCACTACTGAGG 53

DB 228 GCTGAGATAGGAACCCGGCATGAGACTCACCAGACTATGTGG 179

Search completed: May 11, 2003, 04:54:08
Job time : 1049.89 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 64.3484 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactagatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	65	US-09-963-803-8	Sequence 8, Appli
2	65	100.0	317	US-09-963-803-3	Sequence 3, Appli
3	65	100.0	371	US-09-963-803-5	Sequence 5, Appli
4	58	89.2	348	US-09-963-803-4	Sequence 4, Appli
5	58	89.2	398	US-09-963-803-6	Sequence 6, Appli
6	58	89.2	472	US-09-963-803-25	Sequence 25, Appli
7	58	89.2	541	US-09-963-803-24	Sequence 24, Appli
8	58	89.2	604	US-09-963-803-23	Sequence 23, Appli
9	55	84.6	392	US-09-963-803-21	Sequence 21, Appli
10	55	84.6	393	US-09-963-803-19	Sequence 19, Appli
11	55	84.6	462	US-09-963-803-20	Sequence 20, Appli
12	46	70.8	243	US-09-963-803-22	Sequence 22, Appli
13	46	70.8	2600	US-09-963-803-1	Sequence 1, Appli
14	24.2	37.2	2600	US-09-954-456-1157	Sequence 1157, Ap
15	23.6	36.3	273	US-09-777-564-232	Sequence 232, App
16	23.6	36.3	273	US-09-777-564-232	Sequence 232, App
17	22.6	34.8	459	US-09-864-761-3604	Sequence 3604, Ap
18	22.4	34.5	438	US-09-918-995-4240	Sequence 4240, Ap
19	22.4	34.5	906	US-09-938-842A-2162	Sequence 2162, Ap

20	22.4	34.5	1170	10	US-09-974-300-1705	Sequence 1705, Appl
21	22.4	34.5	1902	9	US-09-938-842A-521	Sequence 521, Appl
22	22.4	34.5	3393	9	US-09-738-626-2262	Sequence 2262, Appl
23	22.4	34.5	3309400	9	US-09-738-626-1	Sequence 1, Appl
24	22.2	34.2	1086	9	US-09-991-496-77	Sequence 77, Appl
25	22.2	34.2	1086	10	US-09-874-923-77	Sequence 77, Appl
26	22.2	34.2	1179	10	US-09-759-097-1	Sequence 1, Appl
27	22.2	34.2	1533	12	US-10-074-547-3	Sequence 3, Appl
28	22.2	34.2	4419	12	US-10-074-547-1	Sequence 1, Appl
29	22.2	33.8	42000	9	US-10-081-563-25	Sequence 25, Appl
30	22.2	33.8	402850	9	US-09-844-553-5	Sequence 5, Appl
31	22.2	33.8	465237	10	US-09-933-267A-1	Sequence 1, Appl
32	21.8	33.5	218	10	US-09-923-876-2430	Sequence 2430, Appl
33	21.8	33.5	1740	9	US-09-922-364A-15	Sequence 15, Appl
34	21.8	33.5	1740	9	US-09-254-590-15	Sequence 15, Appl
35	21.8	33.5	1740	9	US-10-115-695-15	Sequence 15, Appl
36	21.8	33.5	1740	9	US-10-116-561-15	Sequence 15, Appl
37	21.8	33.5	1740	9	US-10-115-571-15	Sequence 15, Appl
38	21.8	33.5	1740	9	US-10-115-415-15	Sequence 15, Appl
39	21.8	33.5	1740	9	US-10-116-260-15	Sequence 15, Appl
40	21.8	33.5	3605	10	US-09-881-457A-3	Sequence 3, Appl
41	21.8	33.5	13473	9	US-09-994-064-1	Sequence 1, Appl
42	21.8	33.5	18913	9	US-09-994-064-59	Sequence 59, Appl
43	21.6	33.2	279	10	US-09-864-761-21997	Sequence 21997, Appl
44	21.6	33.2	315	9	US-09-938-842A-1593	Sequence 1593, Appl
45	21.6	33.2	473	10	US-09-864-761-6211	Sequence 6211, Appl

ALIGNMENTS

RESULT 1
US-09-963-803-8
Sequence 8, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963.803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 65
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: S1
US-09-963-803-8
Query Match 100.0%; Score 65; DB 9; Length 65;
Best Local Similarity 100.0%; Pred. No. 6.4e-16;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 CATGCTGCAGACTAGTATCCCGCATCATGATCATCATGATGAGAGATGAA 60
DB 1 CATGCTGCAGACTAGTATCCCGCATCATGATCATCATGATGAGAGATGAA 60
DB 61 TAGCT 65
DB 61 TAGCT 65
RESULT 2
US-09-963-803-3
Sequence 3, Application US/09963803

Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: FR 99/03925
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1116
NAME/KEY: promoter
LOCATION: (1)..(317)
OTHER INFORMATION:
US-09-963-803-3

Query Match 100.0%; Score 65; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGAA 60
DB 8 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGAA 67

QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 3
US-09-963-803-5
Sequence 5; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: FR 99/03925
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 371
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1146
NAME/KEY: promoter
LOCATION: (1)..(371)
OTHER INFORMATION:
US-09-963-803-5

Query Match 100.0%; Score 65; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGAA 60

DB 8 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGAA 67
QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 4
US-09-963-803-4
Sequence 4; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: FR 99/03925
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 348
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1117
NAME/KEY: promoter
LOCATION: (1)..(348)
OTHER INFORMATION:
US-09-963-803-4

Query Match 89.2%; Score 58; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.7e-13;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGATAGCT 65
DB 49 CAGACTAGTATCCGCCGTCAATGACATCATCATCAGTACTGAGAGATGATAGCT 106

RESULT 5
US-09-963-803-6
Sequence 6; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: FR 99/03925
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 398
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1147
NAME/KEY: promoter
LOCATION: (1)..(398)
OTHER INFORMATION:

US-09-963-803-6

Query Match 89.2%; Score 58; DB 9; Length 398;

Best Local Similarity 100.0%; Pred. No. 5.9e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 6

US-09-963-803-25

Sequence 25, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 25

LENGTH: 472

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1169

NAME/KEY: promoter

LOCATION: (1)..(472)

OTHER INFORMATION:

US-09-963-803-25

Query Match 89.2%; Score 58; DB 9; Length 472;

Best Local Similarity 100.0%; Pred. No. 6.2e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 7

US-09-963-803-24

Sequence 24, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 24

LENGTH: 541

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1168

NAME/KEY: promoter

LOCATION: (1)..(541)

OTHER INFORMATION:

US-09-963-803-24

Query Match 89.2%; Score 58; DB 9; Length 541;

Best Local Similarity 100.0%; Pred. No. 6.5e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 8

US-09-963-803-23

Sequence 23, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 23

LENGTH: 604

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1167

NAME/KEY: promoter

LOCATION: (1)..(604)

OTHER INFORMATION:

US-09-963-803-23

Query Match 89.2%; Score 58; DB 9; Length 604;

Best Local Similarity 100.0%; Pred. No. 6.7e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 9

US-09-963-803-21

Sequence 21, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 21

LENGTH: 392

TYPE: DNA

ORGANISM: Artificial Sequence

```
FEATURE:
OTHER INFORMATION: promoter MP-1164
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(392)
OTHER INFORMATION:
US-09-963-803-21
```

```
Query Match      84.6%; Score 55; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 8,7e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 11 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74
```

RESULT 10

```
US-09-963-803-19
Sequence 19, Application US/09963803
Publication No. US20030028922A1
```

```
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP-1162
NAME/KEY: promoter
LOCATION: (1)..(393)
OTHER INFORMATION:
US-09-963-803-19
```

```
Query Match      84.6%; Score 55; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 8,7e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 11 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74
```

RESULT 11

```
US-09-963-803-20
Sequence 20, Application US/09963803
Publication No. US20030028922A1
```

```
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
```

```
LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP-1163
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(462)
OTHER INFORMATION:
US-09-963-803-20
```

```
Query Match      84.6%; Score 55; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 9,1e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 11 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74
```

RESULT 12

```
US-09-963-803-22
Sequence 22, Application US/09963803
Publication No. US20030028922A1
```

```
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 600
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP-1165
NAME/KEY: promoter
LOCATION: (1)..(600)
OTHER INFORMATION:
US-09-963-803-22
```

```
Query Match      84.6%; Score 55; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 9,9e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 11 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCGCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74
```

RESULT 13

```
US-09-963-803-1
Sequence 1, Application US/09963803
Publication No. US20030028922A1
```

```
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
```


NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 243
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: 243 bp Fragment from the intergenic region of commelina yellow mc
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(243)
OTHER INFORMATION:
US-09-963-803-1

Query Match 70.8%; Score 46; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ATCCGCGTCATCATGATCATGACAGTGTGAGGATGATGA 62
DB 1 ATCCGCGTCATCATGATCATGACAGTGTGAGGATGATGA 46

RESULT 14
US-09-954-456-1157/c
Sequence 1157, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
TITLE OF INVENTION: Sets
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1157
LENGTH: 2600
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1157

Query Match 37.2%; Score 24.2; DB 10; Length 2600;
Best Local Similarity 62.3%; Pred. No. 15;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CATGCTGAGAGTAGTATCCGCGTCATCATGATCATGACAGTGTGAGGATGAA 60
DB 676 CTTTCTTCAGAGATGCTCTCCATCATGATCATGATGATCATCATTAATTTGACCAAGGCA 617
QY 61 T 61

DB 616 T 616

RESULT 15
US-10-015-219-232
Sequence 232, Application US/10015219
Publication No. US20030008299A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.493c1
CURRENT APPLICATION NUMBER: US/10/015,219
CURRENT FILING DATE: 2002-03-02
NUMBER OF SEQ ID NOS: 1739
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 232
LENGTH: 273
TYPE: DNA
ORGANISM: Homo sapiens
US-10-015-219-232

Query Match 36.3%; Score 23.6; DB 9; Length 273;
Best Local Similarity 86.7%; Pred. No. 12;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 26 CATCAATGATCATCATGACAGTGTGAGGAG 55
DB 138 CATGAGAGACATCATGACAGTGTGAGGAG 167

Search completed: May 11, 2003, 06:11:38
Job time : 67.3484 secs

XX
XX
Ch
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839

PT Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 23; 91pp; English.
XX
CC The present sequence represents a directional deoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 62 BP; 25 A; 4 C; 21 G; 12 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 21; Length 62;
Best Local Similarity 100.0%; Pred. No. 5e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 GAAGATTAAGTCGGTATTGTGAAAGACATGAGACACATGTAAAGTGAATGTA 60
DB 1 GAAGATTAAGTCGGTATTGTGAAAGACATGAGACACATGTAAAGTGAATGTA 60
XX
OY 61 AG 62
DB 61 AG 62
XX
RESULT 2
AAV14022 standard; DNA; 305 BP.
ID AAV14022 standard; DNA; 305 BP.
XX
AC AAV14022;
XX
DT 18-JUN-1998 (first entry)
XX
DE CaMV promoter PC.
XX
KM Cassava vein mosaic virus; CaMV; promoter; cultivated crop;
KM tissue-specific expression control; transgenic plant; ss.
XX
OS Cassava vein mosaic virus.
XX
PN WO9748819-A1.
XX
PD 24-DEC-1997.
XX
PF 20-JUN-1997; 97WO-US10376.
XX
PR 20-JUN-1996; 96US-0020129.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
XX
DR WPI; 1998-063157/06.
XX
PT Cassava vein mosaic virus promoter - used to express heterologous
PT DNA sequences for producing transgenic plants having altered
PT phenotype(s)
XX
PS Claim 2; Page 77-78; 115pp; English.
XX
CC This sequence represents a cassava vein mosaic virus promoter, and
CC is a nucleic acid molecule of the invention. The promoter is capable of
CC initiating transcription of an operably linked heterologous nucleic acid
CC sequence in a plant cell. The CaMV promoters are active in both monocot

CC and dicot plant species, and therefore can be readily applied to a
CC variety of cultivated crops. Although generally constitutive, the
CC derivative promoters include promoters that can regulate expression in a
CC tissue-specific manner, and therefore are useful for controlling
CC expression of heterologous genes in a tissue-specific manner. The
CC promoters can be used for producing transgenic plants with an altered
CC phenotype.
XX
SQ Sequence 305 BP; 105 A; 46 C; 66 G; 88 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 19; Length 305;
Best Local Similarity 100.0%; Pred. No. 7e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 GAAGATTAAGTCGGTATTGTGAAAGACATGAGACACATGTAAAGTGAATGTA 60
DB 43 GAAGATTAAGTCGGTATTGTGAAAGACATGAGACACATGTAAAGTGAATGTA 102
XX
OY 61 AG 62
DB 103 AG 104
XX
RESULT 3
AAA96837 standard; DNA; 317 BP.
ID AAA96837 standard; DNA; 317 BP.
XX
AC AAA96837;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1116.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 81; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;
 SQ Query Match 100.0%; Score 62; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred.No. 7e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 60
 DB 133 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 192

QY 61 AG 62
 DB 193 AG 194

RESULT 4
 ID AAA96839 standard; DNA; 371 BP.
 AC AAA96839;
 DT 19-FEB-2001 (first entry)
 DE Nucleotide sequence of chimeric expression promoter MP11146.
 KM Promoter; intergenic region; Commelina yellow mottle virus;
 KM chimeric expression promoter; plant vascular expression promoter;
 KM plant green tissue expression promoter; Cassava vein mosaic virus;
 KM transgenic plant; chimera; ss.
 OS Chimeric - Commelina yellow mottle virus.
 OS Chimeric - Cassava vein mosaic virus.
 PN WO200058485-A1.
 PD 05-OCT-2000.
 PF 29-MAR-2000; 2000WO-IB00370.
 PR 29-MAR-1999; 99FR-0003925.
 PA (MERI-) MERISTEM THERAPEUTICS.
 PI Rance I, Gruber V, Theisen M;
 DR WPI; 2000-647238/62.
 PT Chimeric expression promoter for transgenic plant production, comprises
 PT sequence from promoter comprising vascular expression region replaced
 PT with sequence from promoter comprising green tissue expression region
 PS Claim 5; Page 81; 91pp; English.
 CC The present sequence represents a chimeric promoter of the invention.
 CC The specification describes chimeric expression promoters. These
 CC chimeric promoters comprise a nucleic acid sequence which is derived
 CC from a first plant promoter, in which a plant vascular expression
 CC promoter region is replaced with a nucleic acid sequence derived from
 CC a second plant promoter comprising a plant green tissue expression
 CC promoter region. Preferably, the first plant promoter originates from
 CC Commelina yellow mottle virus, and the second plant promoter originates
 CC from the Cassava vein mosaic virus. Especially, the promoters are
 CC derived from intergenic regions. The chimeric promoters are useful
 CC for producing transgenic plants.
 SQ Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;

Query Match 100.0%; Score 62; DB 21; Length 371;
 Best Local Similarity 100.0%; Pred.No. 7.3e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 60
 DB 187 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 246

QY 61 AG 62
 DB 247 AG 248

RESULT 5
 ID AAV14019 standard; DNA; 392 BP.
 AC AAV14019;
 DT 18-JUN-1998 (first entry)
 DE CasMV promoter CVPL.
 KM Cassava vein mosaic virus; CasMV; promoter; cultivated crop;
 KM tissue-specific expression control; transgenic plant; ss.
 OS Cassava vein mosaic virus.
 PN WO9748819-A1.
 PD 24-DEC-1997.
 PF 20-JUN-1997; 97WO-US10376.
 PR 20-JUN-1996; 96US-0020129.
 PA (SCRI) SCRIPPS RES INST.
 PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
 DR WPI; 1998-063157/06.
 PT Cassava vein mosaic virus promoter - used to express heterologous
 PT DNA sequences for producing transgenic plants having altered
 PT phenotype(s)
 PS Claim 2; Page 74; 115pp; English.
 CC This sequence represents a cassava vein mosaic virus promoter, and
 CC is a nucleic acid molecule of the invention. The promoter is capable of
 CC initiating transcription of an operably linked heterologous nucleic acid
 CC sequence in a plant cell. The CasMV promoters are active in both monocot
 CC and dicot plant species, and therefore can be readily applied to a
 CC variety of cultivated crops. Although generally constitutive, the
 CC derivative promoters include promoters that can regulate expression in a
 CC tissue-specific manner, and therefore are useful for controlling
 CC expression of heterologous genes in a tissue-specific manner. The
 CC promoters can be used for producing transgenic plants with an altered
 CC phenotype.
 SQ Sequence 392 BP; 154 A; 64 C; 83 G; 91 T; 0 other;

Query Match 100.0%; Score 62; DB 19; Length 392;
 Best Local Similarity 100.0%; Pred.No. 7.3e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 60
 DB 185 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 244

QY 61 AG 62
 DB 245 AG 246

RESULT 6
 ID AAV14019 standard; DNA; 392 BP.
 AC AAV14019;
 DT 18-JUN-1998 (first entry)
 DE CasMV promoter CVPL.
 KM Cassava vein mosaic virus; CasMV; promoter; cultivated crop;
 KM tissue-specific expression control; transgenic plant; ss.
 OS Cassava vein mosaic virus.
 PN WO9748819-A1.
 PD 24-DEC-1997.
 PF 20-JUN-1997; 97WO-US10376.
 PR 20-JUN-1996; 96US-0020129.
 PA (SCRI) SCRIPPS RES INST.
 PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
 DR WPI; 1998-063157/06.
 PT Cassava vein mosaic virus promoter - used to express heterologous
 PT DNA sequences for producing transgenic plants having altered
 PT phenotype(s)
 PS Claim 2; Page 74; 115pp; English.
 CC This sequence represents a cassava vein mosaic virus promoter, and
 CC is a nucleic acid molecule of the invention. The promoter is capable of
 CC initiating transcription of an operably linked heterologous nucleic acid
 CC sequence in a plant cell. The CasMV promoters are active in both monocot
 CC and dicot plant species, and therefore can be readily applied to a
 CC variety of cultivated crops. Although generally constitutive, the
 CC derivative promoters include promoters that can regulate expression in a
 CC tissue-specific manner, and therefore are useful for controlling
 CC expression of heterologous genes in a tissue-specific manner. The
 CC promoters can be used for producing transgenic plants with an altered
 CC phenotype.
 SQ Sequence 392 BP; 154 A; 64 C; 83 G; 91 T; 0 other;

Query Match 100.0%; Score 62; DB 19; Length 392;
 Best Local Similarity 100.0%; Pred.No. 7.3e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 60
 DB 185 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGGAAAATGTA 244

QY 61 AG 62
 DB 245 AG 246

RESULT 6
 ID AAV14019 standard; DNA; 392 BP.
 AC AAV14019;
 DT 18-JUN-1998 (first entry)
 DE CasMV promoter CVPL.
 KM Cassava vein mosaic virus; CasMV; promoter; cultivated crop;
 KM tissue-specific expression control; transgenic plant; ss.
 OS Cassava vein mosaic virus.
 PN WO9748819-A1.
 PD 24-DEC-1997.
 PF 20-JUN-1997; 97WO-US10376.
 PR 20-JUN-1996; 96US-0020129.
 PA (SCRI) SCRIPPS RES INST.
 PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
 DR WPI; 1998-063157/06.
 PT Cassava vein mosaic virus promoter - used to express heterologous
 PT DNA sequences for producing transgenic plants having altered
 PT phenotype(s)
 PS Claim 2; Page 74; 115pp; English.
 CC This sequence represents a cassava vein mosaic virus promoter, and
 CC is a nucleic acid molecule of the invention. The promoter is capable of
 CC initiating transcription of an operably linked heterologous nucleic acid
 CC sequence in a plant cell. The CasMV promoters are active in both monocot
 CC and dicot plant species, and therefore can be readily applied to a
 CC variety of cultivated crops. Although generally constitutive, the
 CC derivative promoters include promoters that can regulate expression in a
 CC tissue-specific manner, and therefore are useful for controlling
 CC expression of heterologous genes in a tissue-specific manner. The
 CC promoters can be used for producing transgenic plants with an altered
 CC phenotype.
 SQ Sequence 392 BP; 154 A; 64 C; 83 G; 91 T; 0 other;

ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1164.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Cassava vein mosaic virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GAAGATAAGTCGTCGATTGTGAAAGACATAGAGACACATGTAGTGAATGTA 60
DB 135 GAAGATAAGTCGTCGATTGTGAAAGACATAGAGACACATGTAGTGAATGTA 194
XX
QY 61 AG 62
DB 195 AG 196
XX
RESULT 7
ID AAA96853 standard; DNA; 393 BP.
XX
AC AAA96853;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1162.
XX

XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Cassava vein mosaic virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 5; Page 85; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GAAGATAAGTCGTCGATTGTGAAAGACATAGAGACACATGTAGTGAATGTA 60
DB 135 GAAGATAAGTCGTCGATTGTGAAAGACATAGAGACACATGTAGTGAATGTA 194
XX
QY 61 AG 62
DB 195 AG 196
XX
RESULT 8
ID AAV14021 standard; DNA; 411 BP.
XX
AC AAV14021;
XX
DT 18-JUN-1998 (first entry)
XX
DE CeMVV promoter pB.
XX
KM Cassava vein mosaic virus; CeMVV; promoter; cultivated crop;
KM tissue-specific expression control; transgenic plant; ss.
XX
OS Cassava vein mosaic virus.
XX
PN WO9748819-A1.

XX	?
PD	24-DEC-1997.
XX	
PF	20-JUN-1997; 97WO-US10376.
XX	
PR	20-JUN-1996; 96US-0020129.
XX	
PA	(SCRI) SCRIPPS RES INST.
XX	
P1	Beachy RN, De Kochko A, Fauquet C, Verdagner B;
XX	
DR	WPI, 1998-063157/06.
XX	
PT	Casava vein mosaic virus promoter used to express heterologous
PT	DNA sequences for producing transgenic plants having altered
XX	phenotype(s)
XX	
PS	Claim 2; Page 76-77; 115pp; English.
XX	
CC	This sequence represents a casava vein mosaic virus promoter, and
CC	is a nucleic acid molecule of the invention. The promoter is capable of
CC	initiating transcription of an operably linked heterologous nucleic acid
CC	sequence in a plant cell. The CaMV promoters are active in both monocot
CC	and dicot plant species, and therefore can be readily applied to a
CC	variety of cultivated crops. Although generally constitutive, the
CC	derivative promoters include promoters that can regulate expression in a
CC	tissue-specific manner, and therefore are useful for controlling
CC	expression of heterologous genes in a tissue-specific manner. The
CC	promoters can be used for producing transgenic plants with an altered
CC	phenotype.
XX	
SQ	Sequence 411 BP, 157 A, 60 C, 87 G, 107 T; 0 other;
	Query Match 100.0%; Score 62; DB 19; Length 411;
	Best Local Similarity 100.0%; Pred. No. 7.4e-12;
	Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GAAGATTAAGCTCGGTGATTGTGAAGAACAATPAGACACATGTAAAGTGGAAAATGTA 60 DB 149 GAAGATTAAGCTCGGTGATTGTGAAGAACAATPAGACACATGTAAAGTGGAAAATGTA 208
OY	61 AG 62
Db	209 AG 210
	RESULT 9
	AAV14026
ID	AAV14026 standard; DNA; 420 BP.
XX	
AC	AAV14026;
XX	
DT	18-JUN-1998 (first entry)
XX	
DE	CeVMV promoter pDeltaC.
XX	
KM	Casava vein mosaic virus; CeVMV; promoter; cultivated crop;
XX	tissue-specific expression control; transgenic plant; ss.
OS	Casava vein mosaic virus.
XX	
PN	WO9748819-A1.
XX	
PD	24-DEC-1997.
XX	
PF	20-JUN-1997; 97WO-US10376.
XX	
PR	20-JUN-1996; 96US-0020129.
XX	
PA	(SCRI) SCRIPPS RES INST.
XX	
P1	Beachy RN, De Kochko A, Fauquet C, Verdagner B;
XX	

DR WPI; 1998-063157/06.

XX
PT Cassava vein mosaic virus promoter - used to express heterologous
PT DNA sequences for producing transgenic plants having altered
PT phenotype(s)

XX
PS Claim 2; Page 80; 115pp; English.

XX
CC This sequence represents a cassava vein mosaic virus promoter, and
CC is a nucleic acid molecule of the invention. The promoter is capable of
CC initiating transcription of an operably linked heterologous nucleic acid
CC sequence in a plant cell. The CaMV promoters are active in both monocot
CC and dicot plant species, and therefore can be readily applied to a
CC variety of cultivated crops. Although generally constitutive, the
CC derivative promoters include promoters that can regulate expression in
CC tissue-specific manner, and therefore are useful for controlling the
CC expression of heterologous genes in a tissue-specific manner. The
CC promoters can be used for producing transgenic plants with an altered
CC phenotype.

XX
SQ Sequence 420 BP, 149 A, 65 C, 91 G, 115 T, 0 other;

XX
Query Match 100.0%; Score 62; DB 19; Length 420;
Best Local Similarity 100.0%; Pred. No. 7,4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 GAGATGATGGTGGTGGATTTGGAAAGACATAGAGACACATGTAGAGTGAATGTA 60
Db 158 GAAAGATGAGTGGTGGATTTGGAAAGACATAGAGACACATGTAGAGTGAATGTA 217
QY 61 AG 62
Db 218 AG 219

RESULT 10
AAA96854
AAA96854 standard; DNA; 462 BP.
AC
AAA96854;
XX
DT 19-FEB-2001 (first entry)

XX
DE Nucleotide sequence of chimeric expression promoter MP1163.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 86; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX
SQ Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;

Query Match 100.0%; Score 62; DB 21; Length 462;
Best Local Similarity 100.0%; Pred. No. 7.6e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGACATAGACACATGTAGTGAATGTA 60
DB 135 GAAGATTAAGTCGGTGTGTTGTAAGACATAGACACATGTAGTGAATGTA 194

QY 61 AG 62
DB 195 AG 196

RESULT 11
AAV14053
ID AAV14053 standard; DNA; 476 BP.

XX
AC AAV14053;
XX
DT 18-JUN-1998 (first entry)
XX
DE CaMV promoter.
XX
KM Cassava vein mosaic virus; CaMV; promoter; cultivated crop;
KW tissue-specific expression control; transgenic plant; ss.
XX
OS Cassava vein mosaic virus.
XX
PN WO9748819-A1.
XX
PD 24-DEC-1997.
XX
PF 20-JUN-1997; 97WO-US10376.
XX
PR 20-JUN-1996; 96US-0020129.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
XX
PS WPI; 1998-063157/06.
XX
PT Cassava vein mosaic virus promoter - used to express heterologous
PT DNA sequences for producing transgenic plants having altered
PT phenotype(s)
XX
PS Disclosure; Page 87; 115pp; English.

XX This sequence represents a cassava vein mosaic virus promoter, and
CC is a nucleic acid molecule of the invention. The promoter is capable of
CC initiating transcription of an operably linked heterologous nucleic acid
CC sequence in a plant cell. The CaMV promoters are active in both monocot
CC and dicot plant species, and therefore can be readily applied to a
CC variety of cultivated crops. Although generally constitutive, the
CC derivative promoters include promoters that can regulate expression in a
CC tissue-specific manner, and therefore are useful for controlling
CC expression of heterologous genes in a tissue-specific manner. The
CC promoters can be used for producing transgenic plants with an altered

CC phenotype.
XX
SQ Sequence 476 BP; 188 A; 66 C; 110 G; 112 T; 0 other;

Query Match 100.0%; Score 62; DB 19; Length 476;
Best Local Similarity 100.0%; Pred. No. 7.6e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGACATAGACACATGTAGTGAATGTA 60
DB 264 GAAGATTAAGTCGGTGTGTTGTAAGACATAGACACATGTAGTGAATGTA 323

QY 61 AG 62
DB 324 AG 325

RESULT 12
AAA96836
ID AAA96836 standard; DNA; 515 BP.

XX
AC AAA96836;
XX
DT 19-FEB-2001 (first entry)
XX
DE Promoter from intergenic region of Cassava vein mosaic virus.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; ss.
XX
OS Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PS WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 4; Page 80; 91pp; English.

XX The present sequence represents a promoter fragment from the intergenic
CC region of Cassava vein mosaic virus. The promoter is used to construct
CC chimeric expression promoters. These chimeric promoters comprise a
CC nucleic acid sequence which is derived from a first plant promoter,
CC in which a plant vascular expression promoter region is replaced with
CC a nucleic acid sequence derived from a second plant promoter comprising
CC a plant green tissue expression promoter region. Preferably, the first
CC plant promoter originates from Commelina yellow mottle virus, and the
CC second plant promoter originates from the Cassava vein mosaic virus.
CC The chimeric promoters are useful for producing transgenic plants.

XX
SQ Sequence 515 BP; 198 A; 79 C; 109 G; 129 T; 0 other;

Query Match 100.0%; Score 62; DB 21; Length 515;
Best Local Similarity 100.0%; Pred. No. 7.8e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGACATAGACACATGTAGTGAATGTA 60

DB 258 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317
 QY 61 AG 62
 DB 318 AG 319

RESULT 13

AAAF55505
 ID AAF55505 standard; DNA; 515 BP.

AC AAF55505;

DT 29-MAY-2001 (first entry)

XX Nucleotide sequence of a region 5' to the CeMV translation start.

XX CeMV; enhancer domain; enhancer cassette; herbicide resistance;

KM fungal resistance; bacterial disease resistance; insect resistance;

XX plant ripening; plant degradation; plant colour; sweetness; ss.

OS Cassava vein mosaic virus.

PN WO200114573-A1.

PD 01-MAR-2001.

PF 18-AUG-2000; 2000MO-US22595.

PR 19-AUG-1999; 99US-0149763.

XX (PROF-) PROFIGEN INC.

PI Xu D, Nielsen MT;

XX WPI; 2001-211307/21.

PT Enhancer cassette which forms an expression construct together with a promoter useful for manipulating gene expression in plants, comprises

PT duplicated enhancer derived from cassava vein mosaic virus

XX Claim 3; Fig 1; 42pp; English.

CC The present sequence represents a region which is immediately 5' to

CC the Cassava vein mosaic virus (CeMV) translational start site. The

CC sequence is used as an enhancer domain in cassettes of the invention.

CC The specification describes an enhancer cassette comprising a duplicated

CC enhancer derived from CeMV. It can be operably linked to a nucleic

CC acid, the expression of which confers herbicide resistance, fungal,

CC bacterial disease resistance or insect resistance and regulates plant

CC ripening, degradation, colour and sweetness.

XX Sequence 515 BP; 198 A; 78 C; 110 G; 129 T; 0 other;

QY 1 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60

DB 258 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317

QY 61 AG 62

DB 318 AG 319

RESULT 14
 AAV14020
 ID AAV14020 standard; DNA; 524 BP.
 AC AAV14020;

XX 18-JUN-1998 (first entry)
 DT
 XX CeMV promoter CVP2.

XX Cassava vein mosaic virus; CeMV; promoter; cultivated crop;

KM tissue-specific expression control; transgenic plant; ss.

XX Cassava vein mosaic virus.

OS WO9748819-A1.

PN 24-DEC-1997.

PD 20-JUN-1997; 97MO-US10376.

PF 20-JUN-1996; 96US-0020129.

PR (SCRI) SCRIPPS RES INST.

XX Beachy RN, De Kochko A, Fauquet C, Verdaguer B;

XX WPI; 1998-063157/06.

DR Cassava vein mosaic virus promoter - used to express heterologous

XX DNA sequences for producing transgenic plants having altered

XX phenotype(s)

PS Claim 2; Page 75; 115pp; English.

XX This sequence represents a cassava vein mosaic virus promoter, and

CC is a nucleic acid molecule of the invention. The promoter is capable of

CC initiating transcription of an operably linked heterologous nucleic acid

CC sequence in a plant cell. The CeMV promoters are active in both monocot

CC and dicot plant species, and therefore can be readily applied to a

CC variety of cultivated crops. Although generally constitutive, the

CC derivative promoters include promoters that can regulate expression in a

CC tissue-specific manner, and therefore are useful for controlling

CC expression of heterologous genes in a tissue-specific manner. The

CC promoters can be used for producing transgenic plants with an altered

XX phenotype.

XX Sequence 524 BP; 201 A; 79 C; 112 G; 132 T; 0 other;

QY 61 AG 62

DB 322 AG 323

RESULT 15

AAV14018

ID AAV14018 standard; DNA; 526 BP.

AC AAV14018;

XX 18-JUN-1998 (first entry)

DT CeMV promoter pA.

XX Cassava vein mosaic virus; CeMV; promoter; cultivated crop;

XX tissue-specific expression control; transgenic plant; ss.

OS Cassava vein mosaic virus.

XX WO9748819-A1.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 656.672 Seconds
(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62

Sequence: 1 gaagataagtcgctgctg.....tgaagtcggaatgtaag 62

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenBank:*

- 1: gb_da:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sbs:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sbs:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_hcg_mus:*
- 34: em_hcg_pln:*
- 35: em_hcg_rtd:*
- 36: em_hcg_mam:*
- 37: em_hcg_vrt:*
- 38: em_sy:*
- 39: em_hcgo_hum:*
- 40: em_hcgo_mus:*
- 41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	62	6	AX036744 Sequence
2	62	100.0	317	6	AX036737 Sequence
3	62	100.0	371	6	AX036739 Sequence
4	62	100.0	332	6	AX036755 Sequence
5	62	100.0	333	6	AX036753 Sequence
6	62	100.0	462	6	AX036754 Sequence
7	62	100.0	515	6	AX036736 Sequence
8	62	100.0	515	6	AX088388 Sequence
9	62	100.0	532	6	AX202413 Sequence
10	62	100.0	593	6	AX088390 Sequence
11	62	100.0	600	6	AX036756 Sequence
12	62	100.0	838	6	AX014764 Sequence
13	62	100.0	853	6	AX088389 Sequence
14	62	100.0	857	6	AX088391 Sequence
15	62	100.0	931	6	AX088392 Sequence
16	62	100.0	931	6	AX088393 Sequence
17	62	100.0	931	6	AX088393 Sequence
18	62	100.0	1036	6	AX014765 Sequence
19	62	100.0	8158	14	CVU20341 Sequence
20	62	100.0	8159	14	CVU59751 Sequence
21	62	100.0	8340	6	AX329231 Sequence
22	62	100.0	8340	6	AX338536 Sequence
23	62	100.0	9285	6	AX093047 Sequence
24	62	100.0	12241	6	AX412168 Sequence
25	62	100.0	15077	6	AX093052 Sequence
26	50	80.6	301	6	AX036741 Sequence
27	50	80.6	348	6	AX036738 Sequence
28	50	80.6	398	6	AX036740 Sequence
29	50	80.6	472	6	AX036759 Sequence
30	50	80.6	541	6	AX036758 Sequence
31	50	80.6	604	6	AX036757 Sequence
32	30	48.4	192667	2	AL845501 Mus muscu
33	29.2	47.1	192740	2	AL773536 Mus muscu
34	29	46.8	14000	10	AY029613 Mus muscu
35	29	46.8	153483	2	AC068899 Mus muscu
36	29	46.8	155587	2	AC022053 Homo sapi
37	29	46.8	159927	9	AL353590 Human DNA
38	29	46.8	180611	9	AL356284 Human DNA
39	29	46.8	186758	2	AC073678 Mus muscu
40	29	46.8	229700	2	AC073777 Mus muscu
41	28.8	46.5	112537	9	AC112251 Homo sapi
42	28.8	46.5	159173	2	AC117336 Rattus no
43	28.8	46.5	184916	2	AC068998 Mus muscu
44	28.8	46.5	222771	10	AC084069 Mus muscu
45	28	45.2	132164	2	AC115137 Rattus no

ALIGNMENTS

RESULT 1
LOCUS AX036744 62 bp DNA
DEFINITION Sequence 10 from Patent WO0058485.
ACCESSION AX036744
VERSION AX036744.1 GI:11226253
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE
1 (bases 1 to 62)
Rance, I., Theisen, M. and Gruber, V.
AUTHORS Chimeric expression promoters originating from commelina yellow
TITLE mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 10 05-OCT-2000;

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. .62
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S3"
BASE COUNT 25 a 4 c 21 g 12 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA
DB 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
QY 61 AG 62
DB 61 AG 62
RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS AX036737
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 317)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 3 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. .317
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr116"
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 133 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 192
QY 61 AG 62
DB 193 AG 194
RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS AX036739
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 371)

AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 5 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. .371
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1146"
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 187 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 246
QY 61 AG 62
DB 247 AG 248
RESULT 4
AX036755 392 bp DNA linear PAT 16-NOV-2000
LOCUS AX036755
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 392)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow.
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 21 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. .392
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1164"
BASE COUNT 127 a 80 c 87 g 98 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 135 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 194
QY 61 AG 62
DB 195 AG 196
RESULT 5
AX036753 393 bp DNA linear PAT 16-NOV-2000
LOCUS AX036753
DEFINITION Sequence 19 from Patent WO0058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1 (bases 1 to 393)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 19 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

JOURNAL
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1162"

FEATURES
source
promoter 128 a 75 c 93 g 97 t
BASE COUNT 128 a 75 c 93 g 97 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.5e-10; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 60
Db 135 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 194

Qy 61 AG 62
Db 195 AG 196

RESULT 6
AX036754
LOCUS AX036754 462 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 20 from Patent WO0058485.
ACCESSION AX036754
VERSION AX036754.1 GI:11226263
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1 (bases 1 to 462)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 20 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

JOURNAL
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .462
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1163"

FEATURES
source
promoter 148 a 87 c 111 g 116 t
BASE COUNT 148 a 87 c 111 g 116 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.5e-10; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATTAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 60
Db 135 GAAATTAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 194

Qy 61 AG 62
Db 195 AG 196

RESULT 7

AX036736
LOCUS AX036736 515 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 2 from Patent WO0058485.
ACCESSION AX036736
VERSION AX036736.1 GI:11226245
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1 (bases 1 to 515)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 2 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

JOURNAL
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .515
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter from the intergenic region of Cassava Vein
Mosaic Virus of 515 bp in length EMBL U59751"

FEATURES
source
promoter 198 a 79 c 109 g 129 t
BASE COUNT 198 a 79 c 109 g 129 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.5e-10; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 60
Db 258 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 317

Qy 61 AG 62
Db 318 AG 319

RESULT 8
AX088388
LOCUS AX088388 515 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 1 from Patent WO0114573.
ACCESSION AX088388
VERSION AX088388.1 GI:13397256
KEYWORDS
SOURCE Cassava vein mosaic virus.
ORGANISM Cassava vein mosaic virus
Viruses; Retroid viruses; Caulimoviridae; Cassava vein mosaic-like
viruses.

REFERENCE
1 (bases 1 to 515)
Ku, D. and Nielsen, M.T.
Duplicated cassava vein mosaic virus enhancers and uses thereof
Patent: WO 0114573-A 1 01-MAR-2001;
Profigen Inc. (US)

JOURNAL
Location/Qualifiers
1. .515
/organism="Cassava vein mosaic virus"
/db_xref="taxon:38062"
/note="based on cassava vein mosaic virus"

FEATURES
source
promoter 198 a 78 c 110 g 129 t
BASE COUNT 198 a 78 c 110 g 129 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.5e-10; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 60
Db 258 GAAGATAAGTCGGTGAATTGTGAAGACATAGAGACACATGTAAAGTGAATGTA 317

Qy 61 AG 62

Db 318 AG 319

RESULT 9
LOCUS AX0202413 532 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 1 from Patent WO0152620.
ACCESSION AX0202413
VERSION AX0202413.1 GI:15392159
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 532)
AUTHORS Barbas, C.F., Steege, J.T., Guan, X. and Dalmida, B.
TITLE Methods and compositions to modulate expression in plants
JOURNAL Patent: WO 0152620-A 1 26-JUL-2001;
The Scripps Research Institute (US) ; SYNGENTA AGRICULTURAL
DISCOVERY, INC. (CA)
FEATURES
source location/Qualifiers
1..532
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter CgWV"

BASE COUNT 204 a 82 c 111 g 135 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
|||||
Db 273 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 332
|||||

QY 61 AG 62
Db 333 AG 334

RESULT 10
LOCUS AX088390 593 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114573.
ACCESSION AX088390
VERSION AX088390.1 GI:13397258
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 593)
AUTHORS Xu, D. and Nielsen, M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 3 01-MAR-2001;
Profigen Inc. (US)
FEATURES
source location/Qualifiers
1..593
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 228 a 86 c 135 g 144 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
|||||
Db 258 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 317
|||||

QY 61 AG 62
Db 333 AG 334

Db 318 AG 319

RESULT 11
LOCUS AX036756 600 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 22 from Patent WO0058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 600)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 22 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
source location/Qualifiers
1..600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP-165"

BASE COUNT 188 a 111 c 147 g 154 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
|||||
Db 135 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 194
|||||

QY 61 AG 62
Db 195 AG 196

RESULT 12
LOCUS AX014764 838 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 19 from Patent WO9953053.
ACCESSION AX014764
VERSION AX014764.1 GI:10041035
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 838)
AUTHORS Lamberty, M., Hofmann, J., Bulet, P. and Brookhart, G.L.
TITLE Gene coding for heliomycin and use thereof
JOURNAL Patent: WO 9953053-A 19 21-OCT-1999;
LAMBERTY MIREILLE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE-POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source location/Qualifiers
1..838
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 296 a 134 c 175 g 233 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
|||||

Db 264 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 323
QY 61 AG 62
Db 324 AG 325

RESULT 13

AX088389
LOCUS AX088389 853 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 2 from Patent WO0114573.
ACCESSION AX088389
VERSION AX088389.1 GI:13397257
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 853)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 2 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1..853
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 344 a 126 c 191 g 192 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 853;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
Db 596 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 655

QY 61 AG 62
Db 656 AG 657

RESULT 14

AX088391
LOCUS AX088391 857 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 4 from Patent WO0114573.
ACCESSION AX088391
VERSION AX088391.1 GI:13397259
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 857)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 4 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1..857
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 344 a 128 c 191 g 194 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 857;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||||

Db 522 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 581
QY 61 AG 62
Db 582 AG 583

RESULT 15

AX088392
LOCUS AX088392 931 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 5 from Patent WO0114573.
ACCESSION AX088392
VERSION AX088392.1 GI:13397260
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 931)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 5 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1..931
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 374 a 134 c 216 g 207 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 931;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
Db 596 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 655

QY 61 AG 62
Db 656 AG 657

Search completed: May 11, 2003, 06:04:53
Job time : 658.672 secs

US-08-809-297-44

Query March 40.3%; Score 25; DB 2; Length 629;
Best Local Similarity 69.4%; Pred. No. 3.6;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 13 GGTGTTGTGAAGACATAGACACATGTAGTGAATATGTA 61
DB 1 GCGCATTTGCAAGACACACACGACACGAATGTGATTAACATA 49

RESULT 2

US-09-345-882-1

Sequence 1, Application US/09345882

Patent No. 6399373

GENERAL INFORMATION:

APPLICANT: Bougueteloret, Lydie

TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)

FILE REFERENCE: GENSET.031A

CURRENT APPLICATION NUMBER: US/09/345.882

PRIOR APPLICATION NUMBER: US 60/091.315

PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/111.909

PRIOR FILING DATE: 1998-12-10

NUMBER OF SEQ ID NOS: 140

SOFTWARE: Patent.pm

SEQ ID NO 1

LENGTH: 162450

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 72794

OTHER INFORMATION: 5-124-273 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 88073

OTHER INFORMATION: 5-127-261 : polymorphic base A or C

FEATURE:

NAME/KEY: allele

LOCATION: 90842

OTHER INFORMATION: 99-1437-325 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 93714

OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT

FEATURE:

NAME/KEY: allele

LOCATION: 97132

OTHER INFORMATION: 99-1442-224 : polymorphic base G or T

FEATURE:

NAME/KEY: allele

LOCATION: 97152

OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T

FEATURE:

NAME/KEY: allele

LOCATION: 99098

OTHER INFORMATION: 5-130-257 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 99117

OTHER INFORMATION: 5-130-276 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 103806

OTHER INFORMATION: 5-131-395 : polymorphic base A or T

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 108106

OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 108149

OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT

FEATURE:

NAME/KEY: allele

LOCATION: 108308

OTHER INFORMATION: 5-135-357 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 108471

OTHER INFORMATION: 5-136-174 : polymorphic base C or T

FEATURE:

NAME/KEY: allele

LOCATION: 134134

OTHER INFORMATION: 5-140-120 : polymorphic base C or T

FEATURE:

NAME/KEY: allele

LOCATION: 134362

OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 134374

OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA

FEATURE:

NAME/KEY: allele

LOCATION: 146328

OTHER INFORMATION: 5-143-84 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 146345

OTHER INFORMATION: 5-143-101 : polymorphic base A or C

FEATURE:

NAME/KEY: allele

LOCATION: 150329

OTHER INFORMATION: 5-145-24 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 160031

OTHER INFORMATION: 5-148-352 : polymorphic base G or T

FEATURE:

NAME/KEY: allele

LOCATION: 72771..72817

OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30

FEATURE:

NAME/KEY: allele

LOCATION: 72771..72817

OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID31

FEATURE:

NAME/KEY: allele

LOCATION: 88050..88096

OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52

FEATURE:

NAME/KEY: allele

LOCATION: 90819..90865

OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49

FEATURE:

NAME/KEY: allele

LOCATION: 90819..90865

OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match 40.0%; Score 24.8; DB 4; Length 162450;
Best Local Similarity 61.3%; Pred. No. 16;
Matches 38; Conservative 1; Mismatches 23; Indels 0; Gaps 0;

1 GAAGATAGGTCGTGATTTGAAAGACATAGAGACACATGTAAAGTGAATGTA 60

|||||
Db 99061 GAAGAGAGAGAGAAAGATGATGAAAGAAAGAGAGGRTAATGACAGTGAAGAGATGTA 99120
Qy 61 AG 62
Db 99121 AG 99122

|||||
Db 99061 GAAGAGAGAGAGAAAGATGATGAAAGAAAGAGAGGRTAATGACAGTGAAGAGATGTA 99120
Qy 61 AG 62
Db 99121 AG 99122

RESULT 3
US-08-910-925-2
Sequence 2, Application US/08910925
Patent No. 6162601
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
TITLE OF INVENTION: HUMAN PININ SPLICER VARIANT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,925
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0365 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROTO1
CLONE: 53219
US-08-910-925-2

Query Match 39.4%; Score 24.4; DB 4; Length 2369;
Best Local Similarity 63.8%; Pred. No. 8;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

1 GAAGATAGGTCGTGATTTGAAAGACATAGAGACACATGTAAAGTGAATG 58
Db 949 GAGGTAAAGTCTCAGACGAGAGAAAGATGTGAGAGAGACGTAATCAGACAAATG 1006

RESULT 4
US-08-961-527-207
Sequence 207, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Human Genome Sciences, Inc.
;; STREET: 9410 Key West Avenue
;; CITY: Rockville
;; STATE: Maryland
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/961,527
;; FILING DATE:
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: * Brooke, A. Anders
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PB340P1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 207:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3454 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-08-961-527-207

Query Match 38.7%; Score 24; DB 4; Length 3454;
Best Local Similarity 64.3%; Pred. No. 12;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
Db 2616 GAAGTCAAGTTCAGGATTTATGCAAGAGAGATATGCTTTAATGAAATTGAAGA 2671

;; RESULT 5
;; US-08-182-175A-100
;; Sequence 100, Application US/08182175A
;; Patent No. 5559223
;; GENERAL INFORMATION:
;; APPLICANT: Saverio Carl Falco
;; APPLICANT: Sharon J. Keeler
;; APPLICANT: Janet A. Rice
;; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
;; NUMBER OF SEQUENCES: 113
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: E.I. du Pont de Nemours and Company
;; STREET: 1007 Market Street
;; CITY: Wilmington
;; STATE: Delaware
;; COUNTRY: USA
;; ZIP: 19898
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy Disk
;; COMPUTER: Macintosh
;; OPERATING SYSTEM: Macintosh System, 6.0
;; SOFTWARE: Microsoft Word, 4.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/182,175A
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/743,006
;; FILING DATE: 9 August 1991
;; ATTORNEY/AGENT INFORMATION:
;;

;; NAME: Linda Axamethy Floyd
;; REGISTRATION NUMBER: 33,692
;; REFERENCE/DOCKET NUMBER: BB-1031
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (302) 992-4929
;; TELEFAX: (302) 892-7949
;; TELEX: 835420
;; INFORMATION FOR SEQ ID NO: 100:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 130 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; STRAIN: E. coli
;; CELL TYPE: DH5 alpha
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..116
;; OTHER INFORMATION: /function="synthetic storage protein
;; OTHER INFORMATION: /product="protein"
;; OTHER INFORMATION: /gene="sep"
;; OTHER INFORMATION: /standard_name="SSP-seg5"
;; US-08-182-175A-100

Query Match 38.1%; Score 23.6; DB 1; Length 130;
Best Local Similarity 61.3%; Pred. No. 7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Db 11 GAAGTGAAGAAAGCTCAAGAGAGAAATGCTTAAGTGAAGCAAGAAATGTGAAACTGAA 70

;; RESULT 6
;; PCT-US92-06412-100
;; Sequence 100, Application PC/TUS9206412
;; GENERAL INFORMATION:
;; APPLICANT: Saverio Carl Falco
;; APPLICANT: Sharon J. Keeler
;; APPLICANT: Janet A. Rice
;; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing I
;; NUMBER OF SEQUENCES: 113
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: E.I. du Pont de Nemours and Company
;; STREET: 1007 Market Street
;; CITY: Wilmington
;; STATE: Delaware
;; COUNTRY: USA
;; ZIP: 19898
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: Macintosh
;; OPERATING SYSTEM: Macintosh System, 6.0
;; SOFTWARE: Microsoft Word, 4.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/06412
;; FILING DATE: 19920807
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/743,006
;; FILING DATE: 9 August 1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Linda Axamethy Floyd
;; REGISTRATION NUMBER: 33,692
;; REFERENCE/DOCKET NUMBER: BB-1031
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (302) 992-4929
;;

```

? LENGTH: 340 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? STRAIN: E. coli
? CELL TYPE: DH5 alpha
? IMMEDIATE SOURCE:
? CLONE: segment 534 [seg 534]
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3..326
? OTHER INFORMATION: /function= "synthetic seed storage protein"
? OTHER INFORMATION: /product= "protein"
? OTHER INFORMATION: /gene= "ssp"
? OTHER INFORMATION: /standard_name= "SSP-534"
?
US-08-182-175A-104

```

Query Match	38.1%;	Score 23.6;	DB 1;	Length 340;
Best Local Similarity	61.3%;	Pred. No. 9.7;		
Matches	38;	Conservative	0;	Mismatches 24; Indels 0; Gaps 0;
OY	1	GAAGATAGGTGGCTGATTGGAAAGACATAAGAGCACATGTAGTGAATGTA	60	

Qy 1 GAAGATTAAGTCGCGATTGTGTAAAGACACTTAAGACACATTTAAGGTGTGAAATGTGA 60
11 GAAGATTAAGTCGCGATTGTGTAAAGACACTTAAGACACATTTAAGGTGTGAAATGTGA 70
Db 61 AG 62

Db 71 ÅG 72

```

PCT-US92-06412-104
: Sequence 104, Application PC/TUS9206412
: GENERAL INFORMATION:
: APPLICANT: Saverio Carl Falco
: APPLICANT: Sharon J. Keeler
: APPLICANT: Janet A. Rice
: TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing
: NUMBER OF SEQUENCES: 113
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E.I. du Pont de Nemours and Company
: STREET: 1007 Market Street
: CITY: Wilmington
: STATE: Delaware
: COUNTRY: USA
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: Macintosh System, 6.0
: SOFTWARE: Microsoft Word, 4.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/06412
: FILING DATE: 19920807
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/743,006
: FILING DATE: 9 August 1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Linda Akamethy Ployd
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: BB-1031
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (302) 992-4929
: TELEFAX: (302) 892-7949
: TELEAX: 835420
: INFORMATION FOR SEQ. ID NO: 104:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 340 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double

```

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: segment 534 (seg 534)
FEATURE:
NAME/KEY: CDS
LOCATION: 3..326
OTHER INFORMATION: /function= "synthetic seed storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "SSP-534"
PCT-US92-06412-104

Query Match 38.1%; Score 23.6; DB 5; Length 340;
Best Local Similarity 61.3%; Pred. No. 9.7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GAAGTAAGTCGGTATTGGAAGACATAGAGACACATGTAGTGAAGTA 60
DB 11 GAAGTGAAGAAAGCTCAAGAGAGAAATGCTTAGTGAAGACGAATGTGAAGTGA 70

QY 61 AG 62
DB 71 AG 72

RESULT 9
US-09-221-017B-687/C
Sequence 687, Application US/09221017B
Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSE, BRUCE C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: PALO ALTO

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: MONROY, GLADYS H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 687:
SEQUENCE CHARACTERISTICS:
LENGTH: 2240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2240
US-09-221-017B-687

Query Match 38.1%; Score 23.6; DB 4; Length 2240;
Best Local Similarity 64.8%; Pred. No. 15;
Matches 35; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 AAGTAAGTCGGTATTGGAAGACATAGAGACACATGTAGTGAAGTA 55
DB 1405 AAAAATGTTGTTATTGCAAGACCGAGACGCTACAGTGTGCGAAA 1352

RESULT 10
US-08-809-297-43

Sequence 43, Application US/08809297

Patent No. 594850

GENERAL INFORMATION:

APPLICANT: ARAKI, SHIGEKI

APPLICANT: TSUCHIYA, YOHICHI

TITLE OF INVENTION: GENETIC VARIETY IDENTIFYING METHOD IN

TITLE OF INVENTION: HOPS

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/809,297

FILING DATE: 06-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/02121

FILING DATE: 26-JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 7-211328

FILING DATE: 28-JUL-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 8-130586

FILING DATE: 30-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24618

REFERENCE/DOCKET NUMBER: 2589-057-0PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 599 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-809-297-43

Query Match 37.7%; Score 23.4; DB 2; Length 599;
Best Local Similarity 67.3%; Pred. No. 13;

Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 13 GGTATTGTGAAGAGACATAGAGACATGTAGTGAAGTGA 61
Db 1 GGCATTCTGCAAGACACACACGACAGCAAGAAATTGATAACATTA 49

RESULT 11

US-08-098-327E-33
Sequence 33, Application US/08098327E

Patent No. 6270771

GENERAL INFORMATION:

APPLICANT: GUERIN-MARCHAND, Claudine

TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE

TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE

TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/098,327E

FILING DATE: 24-NOV-1993

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91 01286

FILING DATE: 05-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 010830-045

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 464 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

PUBLICATION INFORMATION:

DOCUMENT NUMBER: WO 92/13884

PUBLICATION DATE: 20-AUG-1992

US-08-098-327E-33

Query Match 37.4%; Score 23.2; DB 4; Length 464;
Best Local Similarity 61.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 GAAGATAAGTCGCGTATTGTGAAGAGACATAGAGACACATGTAGTGAAGTGA 60
Db 151 GAAGAAAAAGTTGAAGAAAGTGTGAAGAAATGCGAAGAAAGTGTAGAGAAATGTA 210

RESULT 12
US-08-462-625-33

Sequence 33, Application US/08462625
Patent No. 6319502

GENERAL INFORMATION:

APPLICANT: GUERIN-MARCHAND, Claudine

TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE

TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE

TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/462,625

FILING DATE: 05-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/098,327

FILING DATE: 24-NOV-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91 01286

FILING DATE: 05-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 010830-078

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 464 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

PUBLICATION INFORMATION:

DOCUMENT NUMBER: WO 92/13884

PUBLICATION DATE: 20-AUG-1992

US-08-462-625-33

Query Match 37.4%; Score 23.2; DB 4; Length 464;
Best Local Similarity 61.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 GAAGATAAGTCGCGTATTGTGAAGAGACATAGAGACATGTAGTGAAGTGA 60
Db 151 GAAGAAAAAGTTGAAGAAAGTGTGAAGAAATGCGAAGAAAGTGTAGAGAAATGTA 210

RESULT 13
US-09-641-638-575
Sequence 575, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16

```

PRIORITY APPLICATION NUMBER: US 09/502,330
PRIORITY FILING DATE: 2000-02-11
PRIORITY APPLICATION NUMBER: US 60/133,200
PRIORITY FILING DATE: 1999-05-07
PRIORITY APPLICATION NUMBER: US 09/275,267
PRIORITY FILING DATE: 1999-03-23
PRIORITY APPLICATION NUMBER: US 60/119,917
PRIORITY FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 575
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 214
OTHER INFORMATION: 10-234-179 : deletion AA
NAME/KEY: misc_binding
LOCATION: 195..213
OTHER INFORMATION: 10-234-179.misl
NAME/KEY: primer_bind
LOCATION: 36..56
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 346..366
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_feature
LOCATION: 888
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-575

```

```

      37.4%; Score 23.2; DB 4; Length 1001;
      Best Local Similarity 61.7%; Pred. No. 17;
      Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps
      0;

      QY      3 AGATAAAGTCGGTGAATTGTGAAAAGACATAGAGACACATGTAAGTGGAAATGTAAAG 62
      Db      511 AGCTGGGGGAGAGAGAGAGAGAGAGAGAAACAAGTAAACAACCACTCAGTGGAAAATGTAG 570

      RESULT 14
      US-08-973-462-3
      ; Sequence 3, Application US/08973462B
      ; Patent No. 6191270
      ; GENERAL INFORMATION:
      ; APPLICANT: DRULHE, PIERRE
      ; APPLICANT: DAUBERSIES, PIERRE
      ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
      ; FILE REFERENCE: 0660-0125-0 PCT
      ; CURRENT APPLICATION NUMBER: US/08/973, 462B
      ; CURRENT FILING DATE: 1998-02-06
      ; EARLIER APPLICATION NUMBER: PCT/FR96/00894
      ; EARLIER FILING DATE: 1996-06-12
      ; EARLIER APPLICATION NUMBER: FR 95/07007
      ; NUMBER OF SEQ ID NOS: 29
      ; SOFTWARE: PatentIn Ver. 2.0
      ; SEQ ID NO 3
      ; LENGTH: 1891
      ; TYPE: DNA
      ; ORGANISM: P. falciparum
      ; FEATURE:
      ; NAME/KEY: CDS
      ; LOCATION: (2) ..(1891)
      ; DS-08-973-462-3

```

```

Query March 37.4% Score 23.2; DB 4, Length 1891;
Best Local Similarity 61.7%; Pred. No. 20;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

1 GAAGTAAAGGTGGCGATGTTGGAAGAGACATAGAGGACACATGTAGGTGGAANAATGTA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 320 GAAGAAAAAGTTGAAGAAAGTGTAGAAAGAAATGTA 379

RESULT 15
US-08-973-462-2
Sequence 2, Application US/08973462B
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRULIHER, PIERRE
APPLICANT: DAUBERSIES, PIERRE
TITLE OF INVENTION: MALALIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 5361
TYPE: DNA
ORGANISM: P. falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(5361)
US-08-973-462-2

```

Query March 37.4%; Score 23.2; DB 4; Length 5361;
Best Local Similarity 61.7%; Pred. No. 25;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0.
Oy 1 GAAGATAAGTCGCGATGTTGAAAGACATAGAGACACATGTAAAGTCGAAATGTA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 GAGAAAAAGTTGAAGAAAGTGTAGAGAAAAATGCGAAGAAAGTGTAGAGAAAAATGTA 729

```

```

Oy      1 GAAGATAAGGTCGGTCATTTGTGAAAAGACATATGAGGACATGTAAAGTGGAAAATGTA 60
        |||||
Db      670 GAAGAAAACCTGAAGAAAGCTGAGAGAAAATGACGAAGAAAGCTAGAGAAAATGTA 729

```

Search completed: May 11, 2003, 03:07:35
Job time : 53.8622 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 61.3784 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62

Sequence: 1 gaagataagtcggtgctgctg.....tgtaagtcggaatgtaag 62

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	62	9	US-09-963-803-10
2	62	100.0	317	9	US-09-963-803-3
3	62	100.0	371	9	US-09-963-803-5
4	62	100.0	392	9	US-09-963-803-21
5	62	100.0	393	9	US-09-963-803-19
6	62	100.0	462	9	US-09-963-803-20
7	62	100.0	515	9	US-09-963-803-2
8	62	100.0	532	9	US-09-963-803-22
9	62	100.0	600	9	US-09-963-803-22
10	62	100.0	8340	10	US-09-847-057-4
11	62	100.0	8340	10	US-09-874-926-4
12	62	100.0	12241	12	US-10-033-190-5
13	50	80.6	301	9	US-09-963-803-7
14	50	80.6	348	9	US-09-963-803-4
15	50	80.6	398	9	US-09-963-803-6
16	50	80.6	472	9	US-09-963-803-25
17	50	80.6	541	9	US-09-963-803-24
18	50	80.6	604	9	US-09-963-803-23
19	26	41.9	873	10	US-09-867-550-1223

20	26	41.9	939	9	US-09-860-670-34	Sequence 34, Appl
21	26	41.9	939	10	US-09-764-853-115	Sequence 115, Appl
22	26	41.9	990	9	US-10-160-501-12	Sequence 12, Appl
23	26	41.9	990	10	US-09-870-130-3	Sequence 3, Appl
24	26	41.9	1498	9	US-10-160-501-10	Sequence 10, Appl
25	26	41.9	1498	10	US-09-870-130-1	Sequence 1, Appl
26	26	41.9	1517	10	US-09-834-975-999	Sequence 999, Appl
27	25.4	41.0	672	10	US-09-770-149-403	Sequence 403, Appl
28	25.4	41.0	952	9	US-10-278-173-119	Sequence 119, Appl
29	25.2	40.6	3593	9	US-10-091-504-1634	Sequence 1634, Appl
30	25.2	40.6	3593	10	US-09-764-863-1634	Sequence 1634, Appl
31	24.6	39.7	2444	10	US-09-880-107-3030	Sequence 3030, Appl
32	24.4	39.4	1113	10	US-09-925-300-416	Sequence 416, Appl
33	24.4	39.4	1800	10	US-09-822-849A-1	Sequence 1, Appl
34	24	38.7	241	10	US-09-878-574-13995	Sequence 13995, A
35	24	38.7	422	9	US-09-818-995-17393	Sequence 17393, A
36	23.8	38.4	532	9	US-09-818-995-28694	Sequence 28694, A
37	23.6	38.1	530	10	US-09-764-898-59	Sequence 59, Appl
38	23.6	38.1	728	10	US-09-764-898-109	Sequence 109, Appl
39	23.6	38.1	1201	10	US-09-764-898-131	Sequence 131, Appl
40	23.6	38.1	640681	10	US-09-790-988-1	Sequence 1, Appl
41	23.4	37.7	556	9	US-09-764-891-1302	Sequence 1302, Appl
42	23.4	37.7	850	9	US-09-938-842A-3993	Sequence 3993, Appl
43	23.2	37.4	464	10	US-09-837-344-33	Sequence 33, Appl
44	23.2	37.4	1059	9	US-09-938-842A-2126	Sequence 2126, Appl
45	23.2	37.4	1569	10	US-09-974-300-2651	Sequence 2651, Appl

ALIGNMENTS

RESULT 1
US-09-963-803-10
Sequence 10, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332062
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 62
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Directional desoxynucleotide building block
US-09-963-803-10

Query Match 100.0%; Score 62; DB 9; Length 62;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGGAAGACATGAGACATGTAAGTGGGAAGAATGTA 60
DB 1 GAAGATTAAGTCGGTGTGTTGGAAGACATGAGACATGTAAGTGGGAAGAATGTA 60

QY 61 AG 62
DB 61 AG 62

RESULT 2
US-09-963-803-3
Sequence 3, Application US/09963803.

Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116
NAME/KEY: Promoter
LOCATION: (1)..(317)
OTHER INFORMATION:
US-09-963-803-3

Query Match 100.0%; Score 62; DB 9; Length 317;

Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 133 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 192

QY 61 AG 62
DB 193 AG 194

RESULT 3

US-09-963-803-5
Sequence 5, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 371
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116
NAME/KEY: Promoter
LOCATION: (1)..(371)
OTHER INFORMATION:
US-09-963-803-5

Query Match 100.0%; Score 62; DB 9; Length 371;

Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60

DB 187 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 246

QY 61 AG 62
DB 247 AG 248

RESULT 4

US-09-963-803-21
Sequence 21, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 392
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116
NAME/KEY: Promoter
LOCATION: (1)..(392)
OTHER INFORMATION:
US-09-963-803-21

Query Match 100.0%; Score 62; DB 9; Length 392;

Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 135 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 194

QY 61 AG 62
DB 195 AG 196

RESULT 5

US-09-963-803-19
Sequence 19, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIORITY APPLICATION NUMBER: US/09/963,803
PRIORITY FILING DATE: 1999-03-29
PRIORITY APPLICATION NUMBER: PCT IB00/00370
PRIORITY FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116

Query Match 100.0%; Score 62; DB 9; Length 393;

Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60

FEATURE:
NAME/KEY: promoter
LOCATION: (1) -(393)
OTHER INFORMATION:
US-09-963-803-19

Query Match 100.0%; Score 62; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 135 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 194

QY 61 AG 62
||
DB 195 AG 196

RESULT 6
US-09-963-803-20
Sequence 20, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow m
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925;
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1163
NAME/KEY: promoter
LOCATION: (1) -(462)
OTHER INFORMATION:
US-09-963-803-20

Query Match 100.0%; Score 62; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 135 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 194

QY 61 AG 62
||
DB 195 AG 196

RESULT 7
US-09-963-803-2
Sequence 2, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow m
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 515
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter from the intergenic region of Cassava Vein Mosaic virus
OTHER INFORMATION: of 515 bp in length EMBL
FEATURE:
NAME/KEY: promoter
LOCATION: (1) -(515)
OTHER INFORMATION:
US-09-963-803-2

Query Match 100.0%; Score 62; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 258 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317

QY 61 AG 62
||
DB 318 AG 319

RESULT 8
US-09-765-555-1
Sequence 1, Application US/09765555
Publication No. US20030037355A1
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: Methods and compositions to modulate
FILE REFERENCE: 27801-20014.40
CURRENT APPLICATION NUMBER: US/09/765,555
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 09/620,897
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 60/177,468
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 75
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 532
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter CAVV
US-09-765-555-1

Query Match 100.0%; Score 62; DB 9; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 273 GAAGTAAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 332

QY 61 AG 62
||
DB 333 AG 334

RESULT 9
US-09-963-803-22
Sequence 22, Application US/09963803
Publication No. US20030028922A1

```
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow r
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963.803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 2001-09-26
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP-1165
; NAME/KEY: promoter
; LOCATION: (1) (600)
; OTHER INFORMATION:
US-09-963-803-22
```

```
Query Match          100.0%; Score 62; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 4.5e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
    |||
Db 135 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 194
```

```
QY 61 AG 62
Db 195 AG 196
```

Handwritten: EFT 100

```
RESULT 10
US-09-847-057-4/c
; Sequence 4, Application US/09847057
; Patent No. US20020004943A1
; GENERAL INFORMATION:
; APPLICANT: AGRINOMICS, LLC.
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF A PAGODA PHENOTYPE IN PLAN
; FILE REFERENCE: PAGODA
; CURRENT APPLICATION NUMBER: US/09/847.057
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 8340
; TYPE: DNA
; ORGANISM: Binary vector PAGI4002
US-09-847-057-4
```

```
Query Match          100.0%; Score 62; DB 10; Length 8340;
Best Local Similarity 100.0%; Pred. No. 8.2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
    |||
Db 7811 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 7752
```

```
QY 61 AG 62
Db 7751 AG 7750
```

```
RESULT 11
US-09-874-926-4/c
; Sequence 4, Application US/09874926
; Patent No. US20020010950A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Agrinomics, LLC.
; TITLE OF INVENTION: Identification and characterization of a curly phenotype (cur) i.
; FILE REFERENCE: curly
; CURRENT APPLICATION NUMBER: US/09/874.926
; CURRENT FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 8340
; TYPE: DNA
; ORGANISM: PAGI4002
US-09-874-926-4
```

```
Query Match          100.0%; Score 62; DB 10; Length 8340;
Best Local Similarity 100.0%; Pred. No. 8.2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
    |||
Db 7811 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 7752
```

```
QY 61 AG 62
Db 7751 AG 7750
```

```
RESULT 12
US-10-033-190-5/c
; Sequence 5, Application US/10033190
; Patent No. US20020133848A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Exelixis Plant Sciences, Inc.
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT (AN
; FILE REFERENCE: EP01-002C
; CURRENT APPLICATION NUMBER: US/10/033.190
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/244,685
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 12241
; TYPE: DNA
; ORGANISM: PAG2370
US-10-033-190-5
```

Handwritten: EFT 100

```
Query Match          100.0%; Score 62; DB 12; Length 12241;
Best Local Similarity 100.0%; Pred. No. 9e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
    |||
Db 3217 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAAAGTGGAAAAATGTA 3158
```

```
QY 61 AG 62
Db 3157 AG 3156
```

```
RESULT 13
US-09-963-803-7
; Sequence 7, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963.803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
```

;; PRIOR FILING DATE: 1999-03-29
;; PRIOR APPLICATION NUMBER: PCT IB00/00370
;; PRIOR FILING DATE: 2000-10-05
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 7
;; LENGTH: 301
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Promoter MP1154
;; NAME/KEY: Promoter
;; LOCATION: (1)..(301)
;; OTHER INFORMATION:
US-09-963-803-7

Query Match 80.6%; Score 50; DB 9; Length 301;
Best Local Similarity 98.4%; Pred. No. 4.9e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 60
DB 120 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 178
QY 61 AG 62
DB 179 AG 180

RESULT 14

US-09-963-803-4
;; Sequence 4; Application US/09963803
;; Publication No. US20030028922A1
;; GENERAL INFORMATION:
;; APPLICANT: Meristem Therapeutics
;; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
;; FILE REFERENCE: 184332042
;; CURRENT APPLICATION NUMBER: US/09/963,803
;; PRIOR FILING DATE: 2001-09-26
;; PRIOR APPLICATION NUMBER: FR 99/03925
;; PRIOR FILING DATE: 1999-03-29
;; PRIOR APPLICATION NUMBER: PCT IB00/00370
;; PRIOR FILING DATE: 2000-10-05
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 4
;; LENGTH: 348
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Promoter MP1117
;; NAME/KEY: Promoter
;; LOCATION: (1)..(348)
;; OTHER INFORMATION:
US-09-963-803-4

Query Match 80.6%; Score 50; DB 9; Length 348;
Best Local Similarity 98.4%; Pred. No. 5.1e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 60
DB 167 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 225
QY 61 AG 62
DB 226 AG 227

RESULT 15

US-09-963-803-6
;; Sequence 6; Application US/09963803
;; Publication No. US20030028922A1
;; GENERAL INFORMATION:
;; APPLICANT: Meristem Therapeutics
;; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
;; FILE REFERENCE: 184332042
;; CURRENT APPLICATION NUMBER: US/09/963,803
;; PRIOR FILING DATE: 2001-09-26
;; PRIOR APPLICATION NUMBER: FR 99/03925
;; PRIOR FILING DATE: 1999-03-29
;; PRIOR APPLICATION NUMBER: PCT IB00/00370
;; PRIOR FILING DATE: 2000-10-05
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 6
;; LENGTH: 398
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Promoter MP1147
;; NAME/KEY: Promoter
;; LOCATION: (1)..(398)
;; OTHER INFORMATION:
US-09-963-803-6

Query Match 80.6%; Score 50; DB 9; Length 398;
Best Local Similarity 98.4%; Pred. No. 5.3e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 60
DB 217 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATAATGTA 275
QY 61 AG 62
DB 276 AG 277

Search completed: May 11, 2003, 06:11:41
Job time : 64.3784 secs

Source

1..699

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/clone="nbx0005D02x"

/clone_1lb="CUGI Rice BAC library"

/tissue_type="leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBelOAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Armuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 163 a 250 c 138 g 145 t 3 others

ORIGIN

Query Match 46.8%; Score 29; DB 17; Length 699;

Best Local Similarity 67.2%; Pred. No. 1.1e+02;

Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 GAAGTAAGAGTGGTGCATGTGAAGAGACATAGAGACACATGTAGGTGGAATGTA 60

Db 612 GAAGAGAGATTAGTTAGAGATTAAAGATGTTAGACACCTGTGAAGGGGAAATATTC 671

QY 61 A 61

Db 672 A 672

RESULT 2

RG278522 359 bp mRNA linear EST 21-FEB-2001

LOCUS

DEFINITION

Accession

RG278522

VERSION

RG278522.1 GI:13074882

KEYWORDS

EST.

SOURCE

Neurospora crassa.

ORGANISM

Neurospora crassa

Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 359)

Kupfer,D., Lai,H., Nelson,M. and Roe,B.

ESTs from a Neurospora crassa Sexual cDNA library

Unpublished (2001)

Other_ESTs: as501np.r1

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

Contact Dr. Mary Anne Nelson, Department of Biology, University of New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu) regarding clone availability

Seq primer: M13 Universal Forward Primer

High quality sequence etocp.348.

FEATURES	Location/Qualifiers
SOURCE	1. 359
	/organism="Neurospora crassa"
	/strain="wild type"
	/db_xref="taxon:5141"
	/clone="asc01np"
	/clone_id="Neurospora crassa sexual cDNA library, Uni-zap vector system"
	/issue_type="perithecia (fruiting bodies)"
	/dev_stage="sexual"
	/lab_host="E. coli strain SOLR"
	/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT	182 a 146 c 123 g 97 t
ORIGIN	
Query Match	46.5%; Score 28.8; DB 12; Length 359;
Best Local Similarity	75.0%; Pred. No. 1e+02;
Matches	36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Db	224 ACACGGCAGTGACGAGAGAAATTAAGAGACAGTTAAGCGG 271
Qy	5 ATAACTGCTGATTGTGAAGACATGAGACACATGTAAGCTG 52
RESULT 3	
LOCUS	BG279526 548 bp mRNA linear EST 21-FEB-2001
DEFINITION	b3g07np.f1 Neurospora crassa sexual cDNA library, Uni-zap vector
VERSION	system Neurospora crassa cDNA clone b3g07np 3', mRNA sequence.
KEYWORDS	BG279526
SOURCE	BG279526.1 GI:13076980
ORGANISM	Neurospora crassa.
	Neurospora crassa
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
	Sordariales; Sordariaceae; Neurospora.
REFERENCE	1 (bases 1 to 548)
AUTHORS	Kupfer, D., Lai, H., Nelson, M. and Roe, B.
TITLE	ESTs from a Neurospora crassa Sexual cDNA Library
JOURNAL	Unpublished (2001)
COMMENT	Other ESTs: b3g07np.r1
	Contact: Bruce A. Roe, University of Oklahoma, broeou@ou.edu
	Department of Chemistry and Biochemistry
	Advanced Center for Genome Technology, University of Oklahoma
	620 Parrington Oval, Norman, OK 73019, USA
	Tel: 405 325 4912
	Fax: 405 325 7762
	Email: broeou@ou.edu
	Contact Dr. Mary Anne Nelson, Department of Biology, University of
	New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
	regarding clone availability
	Seq primer: M13 Universal Forward Primer
	High quality sequence stop: 442.
FEATURES	Location/Qualifiers
SOURCE	1. 548
	/organism="Neurospora crassa"
	/strain="wild type"
	/db_xref="taxon:5141"
	/clone="b3g07np"
	/clone_id="Neurospora crassa sexual cDNA library, Uni-zap vector system"
	/issue_type="perithecia (fruiting bodies)"
	/dev_stage="sexual"
	/lab_host="E. coli strain SOLR"
	/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT	182 a 146 c 123 g 97 t
ORIGIN	
Query Match	46.5%; Score 28.8; DB 12; Length 548;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGTATTGTAAGAGACATAGAGACATGTAAGTGG 52
DB 234 ACACGGCGCATGTACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 281

RESULT 4
AI399593

LOCUS AI399593 576 bp mRNA linear EST 08-FEB-1999
DEFINITION NSP6677 Subtracted Perithecial Neurospora crassa cDNA clone SP666

ACCESSION AI399593
VERSION AI399593
KEYWORDS GI:4242680

SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 576)
Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,
Cushing, T., Eretic, A., Fleharty, M., Gorman, M., Judson, K., Miller, R.,
Ortega, J., Pavlova, I., Pera, J., Todisco, S., Trujillo, R.,
Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig,
D.O.

TITLE Expressed sequences from conidial, mycelial, and sexual stages of
Neurospora crassa
JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)
MEDLINE 97435549
COMMENT Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Caretter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngp@biology.unm.edu

FEATURES
source location/Qualifiers

1..576
/organism="Neurospora crassa"
/strain="fl a (FGSC 4347)"
/db_xref="taxon:5141"
/clone="SP666"
/clone_lib="Subtracted Perithecial"
/sex="Mating type a (fluffy), fertilized"
/tissue_type="Perithecia"
/dev_stage="Fruiting Body"
/lab_host="E. coli"
/note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
XhoI; mRNA isolated from 5 day old perithecia (fruiting
bodies) of the fluffy strain fl a (Mating type a),
fertilized with conidia from 74-OR23-IV A (Mating type
A). cDNA directionally cloned into pBluescript SK(-)
using the Uni-ZAP XR vector system (Stratagene, La Jolla,
CA). Previously identified highly expressed clones were
subtracted from this library."

BASE COUNT 168 a 144 c 125 g 139 t
ORIGIN

Query Match 46.5%; Score 28.8; DB 9; Length 576;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGTATTGTAAGAGACATAGAGACATGTAAGTGG 52
DB 101 ACACGGCGCATGTACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 148

RESULT 5
BG280606 585 bp mRNA linear EST 21-FEB-2001
LOCUS BG280606
DEFINITION CSB03mp.fl Neurospora crassa sexual cDNA library, uni-zap vector

ACCESSION BG280606
VERSION BG280606.1 GI:13079182
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 585)
Kupfer, D., Lai, H., Nelson, M. and Roe, B.
ESTs from a Neurospora crassa Sexual cDNA Library
Unpublished (2001)
Other ESTs: CSB03mp.r1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Farrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu

FEATURES
source location/Qualifiers
1..585
/organism="Neurospora crassa"
/strain="wild type"
/db_xref="taxon:5141"
/clone="CSB03mp"
/clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
vector system"
/tissue_type="perithecia (fruiting bodies)"
/dev_stage="sexual"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 187 a 161 c 126 g 111 t
ORIGIN

Query Match 46.5%; Score 28.8; DB 12; Length 585;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGTATTGTAAGAGACATAGAGACATGTAAGTGG 52
DB 224 ACACGGCGCATGTACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 271

RESULT 6
AZ484766/c 214 bp DNA linear GSS 05-OCT-2000
LOCUS AZ484766
DEFINITION IM0311P23F Mouse 10kb plasmid UGCM library Mus musculus genomic
clone UGCM0311P23 F, DNA sequence.

ACCESSION AZ484766
VERSION AZ484766.1 GI:10649926
KEYWORDS GSS.

SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE 1 (bases 1 to 214)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D. Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0311 row: P column: 23
Seq primer: CGTTGTAACAGACGCCACGT
Class: plasmid ends
High quality sequence stop: 214.
Location/Qualifiers:

FEATURES

SOURCE

1. .214
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0311P23"
/clone_1lb="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114[gb|AF129072.1]) a copy-number
inductible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT

52 a 51 c 39 g 72 t

ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 214;
Best Local Similarity 71.2%; Pred. No. 1.5e+02;
Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 2 AAGATACGTCGCTGATTGTGAAGAAGACATGAGACACATGTAAGTGA 53

Db 103 AAGACAAAGTACGTCTGAGGAGAGACCTTGAAGACACATGATTTTGA 52

RESULT 7

LOCUS

AQ165808 383 bp DNA linear GSS 16-OCT-1998

DEFINITION

HS_3071_A1_G12_MR_CIT Approved Human Genomic Sperm Library D Homo

ACCESSION

AQ165808
AQ165808.1 GI:3564003

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 383)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3071 row: M column: 23
Class: BAC ends
High quality sequence stop: 383.
Location/Qualifiers:

FEATURES

SOURCE

1. .383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3071 Col=23 Row=M"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in
E-Coli DH10B"

BASE COUNT

92 a 98 c 86 g 105 t 2 others

ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 383;
Best Local Similarity 71.2%; Pred. No. 1.5e+02;
Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 7 AAGTCGCTGATTGTGAAGAAGACATGAGACACATGTAAGTGAATG 58

Db 169 ATGAAAGTACGTCTGTAAGGAGAGACACACATGCTGTGCAAAATG 118

RESULT 8

LOCUS

AQ212331 418 bp DNA linear GSS 18-SEP-1998

DEFINITION

HS_2240_A2_G10_MR_CIT Approved Human Genomic Sperm Library D Homo

ACCESSION

AQ212331
AQ212331.1 GI:3623532

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 418)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2240 row: M column: 20
Class: BAC ends
High quality sequence stop: 418.
Location/Qualifiers:

FEATURES

SOURCE

1. .418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2240 Col=20 Row=M"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 129 a 63 c 102 g 124 t

ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 418;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 GAAGTAAAGCGGTGATTGTGAAGAAGACATAGAGACATGTAAAGTGAAGAAATTA 60
 DB 176 GGAGATAAATTTGGCAGATTGTGAATATAGACAGTGGAGCATGAAAGAGAGAGTA 235

RESULT 9
 BG279968 407 bp mRNA linear EST 21-FEB-2001
 LOCUS b7h08np.f1 Neurospora crassa sexual cDNA library, Uni-zap vector
 DEFINITION system Neurospora crassa cDNA clone b7h08np 3', mRNA sequence.
 ACCESSION BG279968
 VERSION BG279968.1 GI:13077915
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.

REFERENCE
 1 (bases 1 to 407)
 AUTHORS Kupfer, D., Lai, H., Nelson, M. and Roe, B.
 TITLES ESTs from a Neurospora crassa Sexual cDNA Library
 JOURNAL Unpublished (2001)
 COMMENT Other ESTs: b7h08np.r1
 Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broeou.edu
 Contact Dr. Mary Anne Nelson, Department of Biology, University of
 New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
 regarding clone availability
 Seq primer: M13 Universal Forward Primer
 High quality sequence stop: 402.

FEATURES
 source
 1..407
 /organism="Neurospora crassa"
 /strain="wild type"
 /db_xref="taxon:5141"
 /clone="b7h08np"
 /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
 vector system"
 /tissue_type="perithecia (fruiting bodies)"
 /dev_stage="sexual"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 ; 3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 136 a 110 c 85 g 76 t

ORIGIN
 Query Match 43.9%; Score 27.2; DB 12; Length 407;
 Best Local Similarity 72.9%; Pred. No. 3.2e+02;
 Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 5 ATAAAGTCGCGTATTGTGAAGAAGACATAGAGACATGTAAAGTGG 52
 DB 73 ACACGGGAGTGACAGAGAAAGAAATTAAGAGACAGTAGGGCGG 120

RESULT 10
 BE728034 670 bp mRNA linear EST 15-SEP-2000
 LOCUS 601561322F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:383092 5',
 DEFINITION mRNA sequence.
 ACCESSION BE728034
 VERSION BE728034.1 GI:10142026
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 670)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>,
 TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph. D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC/DCPD/DPF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LLCMS05 row: 1 column: 05
 High quality sequence stop: 670.

FEATURES
 source
 1..670
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3830932"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 179 a 142 c 189 g 160 t

ORIGIN
 Query Match 43.9%; Score 27.2; DB 12; Length 670;
 Best Local Similarity 67.9%; Pred. No. 3.9e+02;
 Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 3 AGATAAGTCGCGTATTGTGAAGAAGACATAGAGACATGTAAAGTGA 58
 DB 371 AGGAGAAGTTGCAGGTGTGCTAAAGACATAGCCGACACGTAAGTGGAAACTG 426

RESULT 11
 A2570132/c 656 bp DNA linear GSS 15-MAY-2001
 LOCUS 270PvG10 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
 DEFINITION A2570132
 ACCESSION A2570132
 VERSION A2570132.1 GI:13980913
 KEYWORDS GSS.
 SOURCE malaria parasite P. vivax.
 ORGANISM Plasmodium vivax
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE
 1 (bases 1 to 656)
 AUTHORS Carlson, J.M.-R. and Dame, J.B.
 TITLES The Plasmodium vivax and P. berghei gene sequence tag projects
 JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
 COMMENT Contact: Dame JB
 Dept. of Pathobiology, College of Veterinary Medicine
 University of Florida
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: M13(-20) forward
 Class: shotgun.

FEATURES
 source
 1..656
 /organism="Plasmodium vivax"
 /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
 497-598)"

```

170 a      140 c      146 g      139 t      1 others
      /db xref="taxon:5855"
      /clone_lib="Pv MEN #30"
      /dev_stage="asexual blood forms"
      /lab_host="Salmix boliviensis"
      /note="vector: pBluescript SK(+), vector DNA, phagemid
excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidiprep
filter, followed by passage through a column of pre-wet
Whaeman CRII powder (1:2 ratio volume of blood to CRII),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nuclease
in the presence of 44% formamide at 500C as described
(Venrick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
500bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

```

Query Match 43.2% Score 26.8; DB 17; Length 656;
Best Local Similarity 68.5%; Pred. No. 5,1e+02;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

7 AAGTTCGGTGAATTGCAAGACACATAGACACATGTAGTGGTGAATGTA 60
Db 155 AAGTACGCTTTTATAGGAAGACATATAGGAAAACTGAAAGGGGAAAAATTTA 102

RESULT 12	
AQ410127	
LOCUS	AQ410127
DEFINITION	HS_5083_BJ_A05_SP6E_RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=659 Col=9 Row=B, DNA sequence.

ORGANISM	<i>Homio sapiens</i>
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 397)
AUTHORS	Mahabes, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. '96 (17), 9739-9744 (1999)
MEDLINE	99380569
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Peter de Jong (peter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.hnsc.washington.edu>

High quality sequence stop: 397.
Location/Qualifiers

```

source
1. .397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=659 Col=9 Row=B"
/clone_1b="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
148 a 50 c 61 g 138 t
BASE COUNT
ORIGIN

```

Query Match	42.9%	Score 26.6	DB 17	Length 397
Best Local Similarity	66.7%	Pred. No. 4.8e+02		
Matches 38	Conservative 0	Mismatches 19	Indels 0	Gaps 0
QY	5	ATAAGTCGTCGATTTCGAAAAGACATCAGACACATCTTAAAGCTGAAAATGTTAA	61	
DB	316	AGAAAGTCGGTGAAATTTCAGAGAAAAGATGAGGAGAAACGTTAAATGAAAGATTA	372	

RESULT	13
LOCUS	A0620909/c
DEFINITION	A0620909 453 bp DNA linear GSS 16-JUN-1999
ACCESSION	H5_2178_B1_D03_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2178 Col=5 Row=H, A06 sequence.
VERSION	A0620909
KEYWORDS	A0620909.1 GI:5083389 GSS.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1 (bases 1 to 453)	Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.	Sequence-tagged connectors: A sequence approach to mapping and procuring the human genome	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)	99380589	Contact: Mahatras GG, Wallace JC, Hood L

Clones may be purchased from Research Genetics (info@reagen.com).
BAC end Web Server: <http://www.htbc.washington.edu>
Plate: 2178 row: H column: 5
Seq primer: T7
Class: BAC ends
High quality sequence stop: 453.

FEATURES	source	Location/Qualifiers
	1. .453	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="plate=2178 Col=5 Row=H"
		/clone_1kb="CTF Approved Human Genomic Sperm Library D"
		/sex="male"
		/note="Organ: sperm; Vector: pBelosAC1; BAC clones in E-Coli DH10B"
BASE COUNT	141 a	79 g 169 t 5 others
ORIGIN	59 c	

Query Match	42.9%	Score 26.6	DB 17	Length 453
Best Local Similarity	65.5%	Pred. No. 5.1e+02		
Matches 38, Conservative	0	Mismatches 20	Indels 0	Gaps 0

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 667.263 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttccattatgaagcactgtg.....aacacactagagatcccg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: GenDb1:*
2: gb_ba:*
3: gb_hcg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vit:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	63	6	AX036746
2	60	95.2	317	6	AX036737
3	60	95.2	371	6	AX036739
4	60	95.2	393	6	AX036753
5	60	95.2	462	6	AX036754
6	60	95.2	600	6	AX036756
7	56	88.9	9285	6	AX039047
8	56	88.9	15077	6	AX093052
9	50.4	80.0	7489	14	CHVWCG
10	49	77.8	243	6	AX036735
11	49	77.8	332	6	AX036755
12	46	73.0	301	6	AX036741
13	46	73.0	348	6	AX036738
14	46	73.0	398	6	AX036740
15	46	73.0	472	6	AX036759
16	46	73.0	541	6	AX036758
17	46	73.0	604	6	AX036757
18	29.4	46.7	165061	2	AC107531
19	29	46.0	526	17	AG025881
20	29	46.0	148152	2	AC022200
21	29	46.0	153936	9	AP001207
22	29	46.0	154587	9	AP001208
23	28.8	45.7	206370	2	AL805937
24	28	44.4	171777	2	AC073151
25	28	44.4	215049	2	AC122807
26	27.6	43.8	121037	2	AL157891
27	27.6	43.8	128228	9	AL138810
28	27.6	43.8	128228	9	HS298105
29	27.6	43.8	172039	2	AC099141
30	27.6	43.8	176425	2	AC090418
31	27.6	43.8	221619	2	AC122844
32	27.4	43.5	140741	9	HS29882
33	27.4	43.5	223678	10	AC027654
34	27.2	43.2	181350	2	AC116233
35	27	42.9	174044	9	AC090675
36	27	42.9	182261	2	AC026134
37	27	42.9	256841	2	AC127284
38	26.8	42.5	78770	2	AC034230
39	26.8	42.5	112630	2	AC025461
40	26.8	42.5	180789	2	AC024633
41	26.8	42.5	194664	2	AC024627
42	26.8	42.5	195369	9	AC113395
43	26.8	42.5	213985	10	AL591129
44	26.6	42.2	55367	9	AC023905
45	26.6	42.2	185771	2	AC111718

ALIGNMENTS

RESULT 1
AX036746
LOCUS AX036746 63 bp DNA
DEFINITION Sequence 12 from Patent WO0058465.
ACCESSION AX036746
VERSION AX036746.1 GI:11226255
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 63)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058465-A 12 05-OCT-2000;

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
Source Location/Qualifiers
1..63
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="directional deoxynucleotide building block S5"
BASE COUNT 21 a 16 c 10 g 16 t
ORIGIN
Query Match 100.0%; Score 63; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 60
DB 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 60
QY 61 CCG 63
DB 61 CCG 63
RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS AX036737
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 317)
Rance, I., Theisen, M., and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 3 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers
1..317
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1116"
promoter 1..317
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN
Query Match 95.2%; Score 60; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 60
DB 258 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 317
RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS AX036739
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 371)
Rance, I., Theisen, M., and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 5 05-OCT-2000;
JOURNAL

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
Source Location/Qualifiers
1..371
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1146"
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN
Query Match 95.2%; Score 60; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 60
DB 312 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 371
RESULT 4
AX036753 393 bp DNA linear PAT 16-NOV-2000
LOCUS AX036753
DEFINITION Sequence 19 from Patent WO0058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 393)
Rance, I., Theisen, M., and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 19 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers
1..393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1162"
promoter 1..393
BASE COUNT 128 a 75 c 93 g 97 t
ORIGIN
Query Match 95.2%; Score 60; DB 6; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 60
DB 334 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAACCAACACCACTAGAGATCC 393
RESULT 5
AX036754 462 bp DNA linear PAT 16-NOV-2000
LOCUS AX036754
DEFINITION Sequence 20 from Patent WO0058485.
ACCESSION AX036754
VERSION AX036754.1 GI:11226263
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 462)
Rance, I., Theisen, M., and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 20 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers

source 1. .462
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPRI163"

Promoter 1. .462
BASE COUNT 148 a 87 c 111 g 116 t
ORIGIN

Query Match 95.2%; Score 60; DB 6; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGATCC 60
DB 403 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGATCC 462

RESULT 6
AX036756 AX036756 600 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 22 from Patent WO0058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 600)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 22 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1. .600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPRI165"

Promoter 1. .600
BASE COUNT 188 a 111 c 147 g 154 t
ORIGIN

Query Match 95.2%; Score 60; DB 6; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGATCC 60
DB 541 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGATCC 600

RESULT 7
AX093047 AX093047 9285 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 52 from Patent WO0118192.
ACCESSION AX093047
VERSION AX093047.1 GI:13509522
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 9285)
AUTHORS Gruber, V. and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods
for obtaining them
Patent: WO 0118192-A 52 15-MAR-2001;
JOURNAL MERISTEM THERAPEUTICS (FR)
FEATURES
source location/Qualifiers
1. .9285
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PMRT1336"

misc_feature 1
/note="PMRT1336 results from the insertion into PMRT1196
of the promoter MPRI165 isolated from plasmid PMRT1322 as
described in PCR patent application PCT/IB00/00370"

Promoter 1. .462
BASE COUNT 2440 a 2252 c 2506 g 2087 t
ORIGIN

Query Match 88.9%; Score 56; DB 6; Length 9285;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGG 56
DB 6330 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGG 6385

RESULT 8
AX093052 AX093052 15077 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 57 from Patent WO0118192.
ACCESSION AX093052
VERSION AX093052.1 GI:13509527
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 15077)
AUTHORS Gruber, V. and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods
for obtaining them
Patent: WO 0118192-A 57 15-MAR-2001;
JOURNAL MERISTEM THERAPEUTICS (FR)
FEATURES
source location/Qualifiers
1. .15077
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PMRT1342"

misc_feature 1
/note="PMRT1342 results from the replacement of the
expression cassette ep35-gus-polyA35S from PMRT1335 by
the expression cassette U5-gus-polyA35S isolated from
PMRT1336"

Promoter 1. .462
BASE COUNT 3672 a 3892 c 4225 g 3288 t
ORIGIN

Query Match 88.9%; Score 56; DB 6; Length 15077;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGG 56
DB 7369 TTCTTATTAGACCTGTGTAGTACTTGAAGAACCAACCAACCACTAGAGG 7424

RESULT 9
CYMVG CYMVG 7489 bp DNA linear VRL 12-SEP-1993
LOCUS
DEFINITION Commelina yellow mottle virus complete circular genome.
X52938
VERSION X52938.1 GI:59047
KEYWORDS complete genome; protease; reverse transcriptase; ribonuclease H.
SOURCE Commelina yellow mottle virus.
ORGANISM Commelina yellow mottle virus.
viruses; Retroviral viruses; Caulimoviridae; Badnavirus.
REFERENCE 1 (bases 1 to 7489)
AUTHORS Olszewski, N.E.
TITLE Direct Substitution
Submitted (02-MAY-1990) Olszewski N.E., University of Minnesota,
Dept. of Plant Biology, 220 Biological Sciences Center, 1445 Gortner
Ave., St. Paul, MN 55108, USA
JOURNAL (revised by [4])
REMARK 2 (bases 1 to 7489)
REFERENCE Medberry, S.L., Lockhart, B.E. and Olszewski, N.E.
AUTHORS

TITLE Properties of Commelina yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus

JOURNAL Nucleic Acids Res. 18 (18), 5505-5513 (1990)

MEDLINE 91016835

PUBMED 1699203

REFERENCE 3 (bases 1 to 7489)

AUTHORS Oleszewski, N.E.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-1990) Oleszewski N.E., University of Minnesota, Dept of Plant Biology, 220 Biological Sciences Center, 1445 Gortner Ave., St. Paul, MN 55108, USA

REFERENCE 4 (bases 1 to 7489)

AUTHORS Oleszewski, N.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-1990)

COMMENT [3] The extra c at position 2672 permits readthrough to result in one ORF (216 kd protein) rather than two (47 kd ORF and 163 kd ORF).

FEATURES

source 1..7489 location/Qualifiers

1..7489 /organism="Commelina yellow mottle virus"

/db_xref="taxon:10653"

/clone="pCOYMW89 (and pCOYMW89)"

1..23 /product="cRNA-Met"

/note="put. cRNA-Met (put. primer for minus strand synthesis)"

300..302 /note="gac was cga"

/citation=[1]

496..1098 /note="ORF 1"

/codon_start=1

/protein_id="CAA37108.1"

/db_xref="GI:59048"

/translation="MNYMLKSHPTPLGLPYSLDPCFMNOVDQVKOILMLSSA KLSSEVIVFTPEVKINLRDLANIHILAHVALGPFVITLYVIDIIFPLKIOGSO KESSENASTLKVQRSLKQIEDLSYQSEBLACRDYISRRPLSKQDVELVY RISQPKFTIKQTBALTEBLKVEEYAKLIHSKGMVLN"

541 /note="c in clone pCOYMW100"

873 /note="t in clone pCOYMW100"

1098..1505 /note="t in clone pCOYMW100"

/note="ORF 2"

/codon_start=1

/protein_id="CAA37109.1"

/db_xref="GI:59049"

/translation="MSNTSKGKYSKALSTNNYLAATIGGATDVGLTGYKQANT TTYLVAKITQIEDLSYQSEBLACRDYISRRPLSKQDVELVY RISQPKFTIKQTBALTEBLKVEEYAKLIHSKGMVLN"

1338..1340 /note="3bp deletion in pCOYMW100"

1506..7166 /note="ORF 3, 216 kd protein"

/codon_start=1

/protein_id="CAA37110.1"

/db_xref="GI:59050"

/translation="MTRRLPAPVOTDGTATSGVVEYEDQIRSRNDORRHITWA GRRRLSLMPGVSSERTIEMOMNPVOQRSMNRHAEVPAEVLKRTGSHVNRV YSHRSEKMMVNGSOVDRSFIOSSSEFVLSRTGIEFTIHGWLVAHQILHRFAGTM ALIVFDTRMSDDRAVLAAMEIDLSBNQIVLVLPIMMTIKSPYHIIQICWTKGSD GWGSDNLITRGLTGLSNTSNVGFAYDKAVANVHLOSGNGVAIGEKEDARFPNG OMNIESKVVVPMQPTEMKAVSNVGTSTSRFENYAAASTSKPQVNEKDEEINBEO EINHNLINDESEIDDEEYIOYORVAAASGVDSGFYDTGVAEIDRCDLDEY VPSETSTPTDSEALIDDEEYAYERKCSDESLOSQPRKYEYTPPOSPHLNE SRSSSSASTSMQDVEEIVRLMKERKKQKKKAQQAQQLSSQADEPTEENIENEN KQADEPTOEIPTHKEPOEIEINIEHVFEEPAFKHLAQLSELVMAAESGGSG VGPQPVNAQPDVNMESPAGYAPATSOATWSNGVNIPIVKSANFRMKPGNQLPSAQ

GDGAMLVFGMNVSPFVPMASITRNYISSFNPDGDKIAMEDLLGEBTERKIFVS
WRMPFDEYONIAKINODGGOALISQIRBIFLGGDPVLGONTVONIAFRKIKOLVC
PNYQSIIRYIMDTVLTAAETGMLSTBEPALISEBFTGMPALIGSRVQAAXYIMPT
SAVNLPSRYFTTNYLTLEQCKASTKRSIALADPCDFTEGIGSGEKRTAKA
TKYTGADHNNHILRYTAKYQKCKCYICQEGHYANQCNKRDQORVAILOSJLKE
NEEVVADKDEEEDEIFSVLGEDEYOEETIMWLEDDIOQIIKEPSKGDLSRRVG
PNPQPAEYQMGVLAKPSKSWRRPIQATLEIINCHNMWTAISTGLACRCKQPLAVO
CHHCAVYQCMCAVHDVQAEIKLSKQVSPSARGKKGAVIDEEDIEGEPILISLO
QENORLOKQVEEIMLREKDALREKDSKASVSTIOSDSEALNIIKEERQ
FKEETMATAOLKEATIVOEEDTIERCMLIBERTENTISATVABEINGLTVNGV
IKPDNEPYINAIIVDTGATACIIQISALPENYBAKATVRSVLAGISTOMTKA
GRLLIGEYFRMPVTVMMNGSLPGIOMTIGSFISLGGIIEEDIIITFYGLVTSI
EISRTTOVANSIEELESDEYINIASVETSPLOEAPARKKIDLKEMKEMKTYGE
NPMPEFKNNKIKCKLNIINPDIKIMGRPIKHTVPGDEAMTQIINLLMGEPILRPS
KHSSTAIVRSGETEIDPIGCKRCKGRVFNKILNENTESQVSLPGINTISYKQ
RSKITKFPULKSGFMQVABESVPTATLAKKLYELVMPFGLKNAFAIRKQND
VFGTEKFLAVYIDILVSETAEQSHQLYTMLOQCKENGILSTPKNIIGPELIDF
LGASLQCTKIKQPHIISKIDFSEDKLTPSGMSMLGILSARVYIDIGLVOPL
ROMKAPYGRMNPETWKNVROIKEKVKULPDLOPKDSFIIETDGCMTGAVCK
WMSKHDPRSTERICAVASGSPNPKSTIDAEIOAIIHGLDKKIYILDKELIIRSD
CEALIKFYKTKTENKRSRVRVLPSPDLGICITTYTPEHIDKHNGLADLSMTNIFI
VERNDSEPYRFTSVEDALVCNDHGRNLISAVINDIITVLR"

1970 /note="a in clone pCOYMW100"

2452 /note="t in clone pCOYMW100"

2672..2674 /note="ccc was cc in [1]"

/citation=[1]

2977 /note="c in clone pCOYMW100"

4611 /note="a in clone pCOYMW100"

4794 /note="t in clone pCOYMW100"

6302 /note="t in clone pCOYMW100"

6434 /note="c in clone pCOYMW100"

7378 /note="c in clone pCOYMW100"

7462 /note="c in clone pCOYMW100"

7478 /note="additional t in pCOYMW100"

7482 /note="c in clone pCOYMW100"

/note="g in clone pCOYMW100"

BASE COUNT 2662 a 1389 c 1577 g 1861 t

ORIGIN

Query Match 80.0%; Score 50.4; DB 14; Length 7489;
Best Local Similarity 90.0%; Pred. No. 1.5e-06;
Matches 54; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TTCCTATTAAACACTGTGTAGTAGCTTAGAAACACACACACACCTAGAGATCC 60
|||||
Db 7318 TTCCTATTAAACACTGTGTAGTAGCTTAGAAACACACACACACACCAATAC 7377

RESULT 10
AX036735
LOCUS AX036735 243 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 1 from Patent WO0058485.
ACCESSION AX036735
VERSION AX036735.1 GI:11226244
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE 1 (bases 1 to 243)
AUTHORS Rance, I., Theissen, M. and Gruber, V.
TITLE Chimic expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus

JOURNAL Patent: WO 0058485-A 1 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .243
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="243 bp promoter fragment from the intergenic region
of Commelina Yellow Mottle Virus"
promoter
1. .243
BASE COUNT 71 a 53 c 45 g 74 t
ORIGIN
Query Match 77.8%; Score 49; DB 6; Length 243;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGCACTGTGTAGTACCTTAGAAAACCAACAACAAC 49
Db 195 TTCCTTATTAGCACTGTGTAGTACCTTAGAAAACCAACAACAAC 243
RESULT 11
LOCUS AX036755 392 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 392)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .392
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1164"
promoter
1. .392
BASE COUNT 127 a 80 c 87 g 98 t
ORIGIN
Query Match 77.8%; Score 49; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 6.9e-06;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGCACTGTGTAGTACCTTAGAAAACCAACAACAAC 49
Db 334 TTCCTTATTAGCACTGTGTAGTACCTTAGAAAACCAACAACAAC 382
RESULT 12
LOCUS AX036741 301 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 7 from Patent WO0058485.
ACCESSION AX036741
VERSION AX036741.1 GI:11226250
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 301)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 7 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES ; GRUBER VERONIQUE (FR)
source
1. .301
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1154"
promoter
1. .301
BASE COUNT 98 a 54 c 74 g 75 t
ORIGIN
Query Match 73.0%; Score 46; DB 6; Length 301;
Best Local Similarity 100.0%; Pred. No. 7.1e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACCTTAGAAAACCAACAACAACCTTAGAGATCC 60
Db 256 ACTGTGTAGTACCTTAGAAAACCAACAACAACCTTAGAGATCC 301
RESULT 13
LOCUS AX036738 348 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 4 from Patent WO0058485.
ACCESSION AX036738
VERSION AX036738.1 GI:11226247
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 4 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .348
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1117"
promoter
1. .348
BASE COUNT 116 a 70 c 78 g 84 t
ORIGIN
Query Match 73.0%; Score 46; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACCTTAGAAAACCAACAACAACCTTAGAGATCC 60
Db 303 ACTGTGTAGTACCTTAGAAAACCAACAACAACCTTAGAGATCC 348
RESULT 14
LOCUS AX036740 398 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 6 from Patent WO0058485.
ACCESSION AX036740
VERSION AX036740.1 GI:11226249
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 398)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 6 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .398
Location/Qualifiers

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1147"
1..398
BASE COUNT 128 a 80 c 93 g 97 t
ORIGIN

Query Match 73.0%; Score 46; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTTGTTAGTAGCTTAGAAAACCAACAACTAGAGATCC 60
DB 353 ACTTGTTAGTAGCTTAGAAAACCAACAACTAGAGATCC 398

RESULT 15
AX036759 472 bp DNA linear PAT 16-NOV-2000
LOCUS AX036759
DEFINITION Sequence 25 from Patent WO0058485.
ACCESSION AX036759
VERSION AX036759.1 GI:11226268
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 472)

REFERENCE
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
JOURNAL motile virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
location/Qualifiers

FEATURES
source 1..472
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1147"
1..472
BASE COUNT 149 a 92 c 112 g 119 t
ORIGIN
promoter

Query Match 73.0%; Score 46; DB 6; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTTGTTAGTAGCTTAGAAAACCAACAACTAGAGATCC 60
DB 427 ACTTGTTAGTAGCTTAGAAAACCAACAACTAGAGATCC 472

Search completed: May 11, 2003, 06:04:58
Job time : 672.263 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 122.526 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-12
Perfect score: 63
Sequence: 1 ttccctatcaagcactgt.....aacacatagagatccccc 63

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	100.0	63	21	AAA96846
2	60	95.2	317	21	AAA96837
3	60	95.2	371	21	AAA96839
4	60	95.2	393	21	AAA96853
5	60	95.2	462	21	AAA96854
6	60	95.2	600	21	AAA96856
7	49	77.8	243	21	AAA96835
8	49	77.8	392	21	AAA96855
9	46	73.0	301	21	AAA96841

10	46	73.0	348	21	AAA96838	Nucleotide sequenc
11	46	73.0	398	21	AAA96840	Nucleotide sequenc
12	46	73.0	472	21	AAA96859	Nucleotide sequenc
13	46	73.0	541	21	AAA96858	Nucleotide sequenc
14	46	73.0	604	21	AAA96857	Nucleotide sequenc
15	27	42.9	2928	23	AAA89944	DNA encoding novel
16	27	42.9	4600	23	AAA868229	DNA encoding novel
17	26.4	41.9	2786	22	AAA14506	Human cDNA sequenc
18	26	41.3	4289	23	ABP03686	Drosophila melanog
19	25.8	41.0	14919	22	AA46505	Tumour suppressor
20	25	39.7	694	21	AA14716	Aspergillus oryzae
21	24.8	39.4	815	21	AA15867	Human prostate can
22	24.8	39.4	1969	22	AAH14864	Human cDNA sequenc
23	24.6	39.0	1196	24	ABL40669	Human Fe-S protein
24	24.4	38.7	2203	22	AA160067	Human polynucleoti
25	24.4	38.7	2218	22	AA158281	Human polynucleoti
26	24.4	38.7	2915	22	AA521348	Human cDNA sequenc
27	24.4	38.7	52562	22	AAK86669	Human immune/haema
28	24.4	38.7	53075	22	AAK86671	Human immune/haema
29	24	38.1	329	22	AA187625	Human polynucleoti
30	24	38.1	777	24	ABN66600	Streptococcus poly
31	24	38.1	1023	22	AAH33092	Human colon cancer
32	24	38.1	1032	20	AAZ7434	Human secreted pro
33	24	38.1	1034	20	AAZ7315	Human secreted pro
34	24	38.1	3245	22	AAH54626	S. epidermidis gen
35	23.8	37.8	10202	23	ABP02571	Drosophila melanog
36	23.8	37.8	10369	22	AA46304	Tumour suppressor
37	23.8	37.8	10369	24	ABJ32393	Human immune syste
38	23.8	37.8	19603	23	ABJ02570	Drosophila melanog
39	23.6	37.5	1203	23	AA580397	Drosophila melanog
40	23.6	37.5	2462	23	ABJ22798	Chemically treated
41	23.6	37.5	5102	24	ABL92274	Drosophila melanog
42	23.4	37.1	6847	23	ABL14139	Streptococcus pneu
43	23.4	37.1	14872	19	AAV52205	Drosophila melanog
44	23.4	37.1	22948	23	ABL14138	Drosophila melanog
45	23.2	36.8	411	23	ABV46444	Human prostate exp

ALIGNMENTS

RESULT 1	
AAA96846	AAA96846 standard; DNA; 63 BP.
XX	XX
AC	AAA96846;
XX	XX
DT	19-FEB-2001 (first entry)
XX	XX
DE	Directional deoxynucleotide building block 55.
XX	XX
KM	Promoter; intergenic region; Commelina yellow mottle virus;
KM	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; ss.
XX	XX
OS	Synthetic.
XX	XX
FN	WO200058485-A1.
XX	XX
PD	05-OCT-2000.
XX	XX
PP	29-MAR-2000; 2000MO-IB00370.
XX	XX
PR	29-MAR-1999; 99FR-0003925.
XX	XX
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	XX
PI	Rance I, Gruber V, Theisen M;
XX	XX
DR	WPI; 2000-647238/62.
XX	XX
PT	Chimeric expression promoter for transgenic plant production, comprises

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;

Query Match 95.2%; Score 60; DB 21; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps

1 TTCTCTATTATTAAGCACTGTGTAGTACTTGAAGAAACCAACAACTAGAGATCC 60
Dh 258 TTCTCTATTATTAAGCACTGTGTAGTACTTGAAGAAACCAACAACTAGAGATCC 317

RESULT 3
AAA96839
ID AAA96839 standard; DNA; 371 BP.
AC AAA96839;
AT 19-FEB-2001 (first entry)
XX
XX Nucleotide sequence of chimeric expression promoter MP1116.
DE
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
PN
PD 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-1B00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
XX sequence from promoter comprising vascular expression region replaced
XX with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 81; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
XX The specification describes chimeric expression promoters. These
XX chimeric promoters comprise a nucleic acid sequence which is derived
XX from a first plant promoter, in which a plant vascular expression
XX promoter region is replaced with a nucleic acid sequence derived from
XX a second plant promoter comprising a plant green tissue expression
XX promoter region. Preferably, the first plant promoter originates from
XX Commelina yellow mottle virus, and the second plant promoter originates
XX from the Cassava vein mosaic virus. Especially, the promoters are
XX derived from intergenic regions. The chimeric promoters are useful
XX for producing transgenic plants.

XX Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;
SQ
Query Match 95.2%; Score 60; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 312 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 371

RESULT 4
ID AAA96853
AAA96853 standard; DNA; 393 BP.
XX
AC AAA96853;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1162.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 85; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;
SO
Query Match 95.2%; Score 60; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 334 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 393

RESULT 5
ID AAA96854
AAA96854 standard; DNA; 462 BP.
XX
AC AAA96854;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1163.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;
SO
Query Match 95.2%; Score 60; DB 21; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 403 TTCCTATTAAAGCACTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 462

RESULT 6
ID AAA96856
AAA96856 standard; DNA; 600 BP.
XX
AC AAA96856;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1165.
XX

KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86-87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 600 BP; 188 A; 111 C; 147 G; 154 T; 0 other;
Query Match 95.2%; Score 60; DB 21; Length 600;
Best Local Similarity 100.0%; Pred. No. 1,2e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTATTAAGCATTGTGTAGTCTTAGAAAAACCAACCAACCTAGAGATCC 60
DB 541 TTCCTATTATTAAGCATTGTGTAGTCTTAGAAAAACCAACCAACCTAGAGATCC 600
RESULT 7
AAA96835
ID AAA96835 standard; DNA; 243 BP.
XX
AC AAA96835;
XX
DT 19-FEB-2001 (first entry)
XX
DE Promoter from intergenic region of Commelina yellow mottle virus.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; ss.
XX
OS Commelina yellow mottle virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX

PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 4; Page 79; 91pp; English.
XX
XX The present sequence represents a promoter fragment from the intergenic
CC region of Commelina yellow mottle virus. The promoter is a strong
CC promoter in vascular and reproductive tissues. The promoter is used to
CC construct chimeric expression promoters. These chimeric promoters
CC comprise a nucleic acid sequence which is derived from a first
CC plant promoter, in which a plant vascular expression promoter region is
CC replaced with a nucleic acid sequence derived from a second plant
CC promoter comprising a plant green tissue expression promoter region.
CC Preferably, the first plant promoter originates from Commelina yellow
CC mottle virus, and the second plant promoter originates from the Cassava
CC vein mosaic virus. The chimeric promoters are useful for producing
CC transgenic plants.
XX
SQ Sequence 243 BP; 71 A; 53 C; 45 G; 74 T; 0 other;
Query Match 77.8%; Score 49; DB 21; Length 243;
Best Local Similarity 100.0%; Pred. No. 7.7e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTATTAAGCATTGTGTAGTCTTAGAAAAACCAACCAAC 49
DB 195 TTCCTATTATTAAGCATTGTGTAGTCTTAGAAAAACCAACCAAC 243
RESULT 8
AAA96855
ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter Mpr164.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX

PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
PS
XX Claim 5, Page 86, 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, the first plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP, 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 77.8%; Score 49; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 8,4e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TTCCTATTATTAAGCACTTGCTAGTACCTTAGAAGAAACCAACACACAC 49
DB 334 TTCCTATTATTAAGCACTTGCTAGTACCTTAGAAGAAACCAACACACAC 382
AAAG6841
ID AAA96841 standard; DNA; 301 BP.
AC
XX AAA96841;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MB1154.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
PD
XX 05-OCT-2000.
PD
XX 29-MAR-2000; 2000WO-1B00370.
PF
XX 29-MAR-1999; 99FR-0003925.
RR
XX (MERI-) MERISTEM THERAPEUTICS.
PA
XX Rance I, Gruber V, Theisen M;
PI
XX WPI: 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
PS
XX Claim 5, Page 82, 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 301 BP; 98 A; 54 C; 74 G; 75 T; 0 other;

5Q

Query Match 73.0%; Score 46; DB 21; Length 301;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 15 ACTGTGTAGTACTTAGAAAACCAACCAACAACCTTAGAGGATCC 60
Db 256 ACTGTGTAGTACTTAGAAAACCAACCAACAACCTTAGAGGATCC 301

RESULT 10
AAA96838
AAA96838 standard; DNA; 348 BP.
AAA96838;
19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1117.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO20058485-A1.
FN
PD 05-OCT-2000.
XX
PP 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
DR
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX Claim 5; Page 81; 91pp; English.
PS
XX

The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

8Q Sequence 348 BP; 116 A; 70 C; 78 G; 84 T; 0 other;

73.0%; Score 46; DB 21; Length 348;
Query Match

Best Local Similarity 100.0%; Pred. No. 9,4e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 60
DB 303 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 348

RESULT 11

AAA96840
ID AAA96840 standard; DNA; 398 BP.

AC AAA96840;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1147.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

PF 29-MAR-2000; 2000WO-1B00370.

PR 29-MAR-1999; 99FR-0003925.

PS (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 82; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

SQ Sequence 398 BP; 128 A; 80 C; 93 G; 97 T; 0 other;

Query Match 73.0%; Score 46; DB 21; Length 398;

Best Local Similarity 100.0%; Pred. No. 9.7e-07;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 60

DB 353 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 398

RESULT 12

AAA96859
ID AAA96859 standard; DNA; 472 BP.

XX AAA96859;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1169.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

PF 29-MAR-2000; 2000WO-1B00370.

PR 29-MAR-1999; 99FR-0003925.

PS (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 88; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

SQ Sequence 472 BP; 149 A; 92 C; 112 G; 119 T; 0 other;

Query Match 73.0%; Score 46; DB 21; Length 472;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 60

DB 427 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCTTAGAGATCC 472

RESULT 13

AAA96858
ID AAA96858 standard; DNA; 541 BP.

AC AAA96858;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1168.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

XX OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX WPI; 2000-647238/62.
DR
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 5; Page 87-88; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 541 BP; 169 A; 104 C; 130 G; 138 T; 0 other;
XX
Query Match 73.0%; Score 46; DB 21; Length 541;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 15 ACTGTGTAGTACTGAGAAACCAACCACTAGAGATCC 60
Db 496 ACTGTGTAGTACTGAGAAACCAACCACTAGAGATCC 541
XX
RESULT 14
ID AAA96857
AC AAA96857 standard; DNA; 604 BP.
XX
AC AAA96857;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1167.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925;
XX

XX PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX WPI; 2000-647238/62.
DR
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 5; Page 87; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
XX
Query Match 73.0%; Score 46; DB 21; Length 604;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 15 ACTGTGTAGTACTGAGAAACCAACCACTAGAGATCC 60
Db 559 ACTGTGTAGTACTGAGAAACCAACCACTAGAGATCC 604
XX
RESULT 15
ID AAS89944
AC AAS89944 standard; cDNA; 2928 BP.
XX
AC AAS89944;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #25748.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG25757.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX

PS Claim 1; SEQ ID No 25748; 103bp; English.

XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcc_sequences.

XX
SQ Sequence 2928 BP; 742 A; 773 C; 737 G; 676 T; 0 other;

Query Match 42.9%; Score 27; DB 23; Length 2928;

Best Local Similarity 85.7%; Pred. No. 7;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 CTTAGAAAACCAACACCACTTAGAGATCCCC 62

DB 903 CTTGCGTAGCCACACACAGCTTAGAGATCCCC 937

Search completed: May 11, 2003, 03:04:03
Job time : 124.526 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 25.2632 Seconds

(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttctctatctaagcactgt.....aacacactagagatcccg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PTUS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	38.1	1032	US-09-227-357-134	Sequence 134, App
2	24	38.1	1034	US-09-227-357-15	Sequence 15, Appl
3	23.4	37.1	14872	US-08-961-527-72	Sequence 37, Appl
4	22.4	35.6	518	US-09-280-116-245	Sequence 245, App
5	22.2	35.2	836	US-08-911-434A-5	Sequence 5, Appl
6	22	34.9	2885	US-08-920-812-4	Sequence 4, Appl
7	22	34.9	2885	US-08-920-827-4	Sequence 4, Appl
8	22	34.9	2885	US-08-921-177-4	Sequence 4, Appl
9	22	34.9	2885	US-08-362-577C-4	Sequence 4, Appl
10	22	34.9	2885	US-08-920-828-4	Sequence 4, Appl
11	21.8	34.6	96	US-08-755-587-4	Sequence 4, Appl
12	21.8	34.6	1056	US-08-755-587-28	Sequence 28, Appl
13	21.8	34.6	1917	US-08-755-587-1	Sequence 1, Appl
14	21.8	34.6	3131	US-08-840-466A-20	Sequence 20, Appl
15	21.8	34.6	3131	US-09-696-188B-20	Sequence 20, Appl
16	21.8	34.6	7240	US-08-755-587-15	Sequence 15, Appl
17	21.8	34.6	11283	US-08-603-753D-3	Sequence 3, Appl
18	21.8	34.6	11283	US-09-099-753-3	Sequence 3, Appl
19	21.8	34.6	11283	US-08-986-106-3	Sequence 3, Appl
20	21.8	34.6	11385	US-08-639-501-1	Sequence 1, Appl
21	21.8	34.6	11385	US-09-044-946-1	Sequence 1, Appl
22	21.8	34.6	11385	US-09-044-908-1	Sequence 1, Appl
23	21.8	34.6	19718	US-08-961-527-99	Sequence 99, Appl
24	21.8	34.6	43360	US-09-453-702B-206	Sequence 206, App
25	21.6	34.6	45325	US-09-453-702B-261	Sequence 261, App
26	21.6	34.3	444	US-07-708-038-3	Sequence 3, Appl
27	21.6	34.3	444	US-08-127-995-3	Sequence 3, Appl

28	21.6	34.3	1900	1	US-08-764-343-2	Sequence 2, Appl
C 29	21.6	34.3	1960	4	US-09-513-057C-12	Sequence 12, Appl
C 30	21.6	34.3	3617	4	US-09-513-057C-14	Sequence 14, Appl
C 31	21.4	34.0	272	4	US-09-537-357-31	Sequence 31, Appl
C 32	21.4	34.0	1715	4	US-09-537-357-56	Sequence 56, Appl
C 33	21.4	34.0	4471	2	US-08-615-942A-1	Sequence 1, Appl
C 34	21.2	33.7	198	4	US-09-134-001C-309	Sequence 309, App
C 35	21.2	33.7	7432	1	US-07-852-260-1	Sequence 1, Appl
C 36	21.2	33.7	7432	2	US-08-461-503-1	Sequence 1, Appl
C 37	21.2	33.7	7432	3	US-08-465-250-1	Sequence 1, Appl
C 38	21.2	33.7	11721	4	US-09-026-039-3	Sequence 3, Appl
C 39	21.2	33.7	11721	4	US-09-026-039-3	Sequence 3, Appl
C 40	21.2	33.7	112132	4	US-09-741-150-3	Sequence 3, Appl
C 41	21	33.3	375	4	US-09-134-001C-2507	Sequence 2507, App
C 42	21	33.3	1029	4	US-09-134-001C-1353	Sequence 1353, App
C 43	21	33.3	1993	1	US-08-487-890A-108	Sequence 108, App
C 44	21	33.3	1993	2	US-08-478-435-108	Sequence 108, App
C 45	21	33.3	1993	2	US-08-337-483-108	Sequence 108, App

ALIGNMENTS

RESULT 1
US-09-227-357-134
; Sequence 134, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13664
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18

PATENTNO: 6342581
 GENERAL INFORMATION:
 APPLICANT: Fischer et al.
 TITLE OF INVENTION: 123 Human Secreted Proteins
 FILE REFERENCE: P20101
 CURRENT APPLICATION NUMBER: US/09/227,357
 CURRENT FILING DATE: 1999-01-08
 EARLIER APPLICATION NUMBER: PCT/US98/13684
 EARLIER FILING DATE: 1998-07-07
 EARLIER APPLICATION NUMBER: 60/051,926
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,793
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,925
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/051,929
 EARLIER FILING DATE: 1997-07-08
 EARLIER APPLICATION NUMBER: 60/052,803
 EARLIER FILING DATE: 1997-07-08

RESULT 3
US-08-961-527-72/c

Sequence 72, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 14872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-72

Query Match 37.1%; Score 23.4; DB 4; Length 14872;
Best Local Similarity 63.2%; Pred. No. 22;
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTCCTATTATGACCTGTGTAGCTTAGAAACCAACACCAACTAGAGGA 57
DB 1596 TCCTGCAITTAACATATGATAGTACCAAGATACCGCAAAACAGCTAGAGGA 1540

RESULT 4
US-09-280-116-245
Sequence 245, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 245
LENGTH: 518
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
US-09-280-116-245

Query Match 35.6%; Score 22.4; DB 4; Length 518;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 7 ATTATGACCTGTGTAGCTTAGAAACCAACACCAACTAGAGATCCCC 62
DB 313 ATTATATCGACCTTGGAGAGACCAAAAAAATTCAACACAAATTTGATTTCCCC 368

RESULT 5
US-08-911-434A-5/c
Sequence 5, Application US/08911434A
Patent No. 5959176
GENERAL INFORMATION:
APPLICANT: TORIKAI, Satomi
APPLICANT: OEDA, Kenji
TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22040
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,434A
FILING DATE: 12-AUG-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Raymond C.
REGISTRATION NUMBER: 21,066
REFERENCE/DOCKET NUMBER: 2185-0199P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)205-8000
TELEFAX: (703)205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Daucus carota L.
INDIVIDUAL ISOLATE: Kuroda Gosun
FEATURE:
NAME/KEY: terminator
LOCATION: 1..836
US-08-911-434A-5

Query Match 35.2%; Score 22.2; DB 2; Length 836;
Best Local Similarity 69.8%; Pred. No. 33;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 AAGACCTTGTTGTAGCTTAGAAACCAACACCAACTAG 53
DB 724 AAGCACTTATATGAACTTAAAGCCCAAAAAGAAAGTTAG 682

RESULT 6
US-08-920-812-4
Sequence 4, Application US/08920812
Patent No. 5763188
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotugu
TITLE OF INVENTION: Eda, Soji
NUMBER OF SEQUENCES: 25

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/920,812
;; FILING DATE: 29-AUG-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/362,577
;; FILING DATE: 27-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rin-Laures, Li-Hsien
;; REGISTRATION NUMBER: 33,547
;; REFERENCE/DOCKET NUMBER: 19036/32420
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2885 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; ORIGINAL SOURCE:
;; ORGANISM: Staphylococcus aureus
;; STRAIN: Clinical isolate SA-77
;; US-08-920-812-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 TTCTTATTAGACCTGTGTAGTACCTTAGAAACCAACACCACTAGA 54
1639 TTGCTGATGACGATTTTATTATTAATGATGATGCCAATATACCACTATA 1632

;; RESULT 7
;; US-08-920-827-4
;; Sequence 4, Application US/08920827
;; Patent No. 570375
;; GENERAL INFORMATION:
;; APPLICANT: Ohno, Tsuneya
;; APPLICANT: Matsuhisa, Akio
;; APPLICANT: Uehara, Hirotsugu
;; APPLICANT: Eda, Soji
;; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/920,827

;; FILING DATE: 29-AUG-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/362,577
;; FILING DATE: 27-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rin-Laures, Li-Hsien
;; REGISTRATION NUMBER: 33,547
;; REFERENCE/DOCKET NUMBER: 19036/32420
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2885 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; ORIGINAL SOURCE:
;; ORGANISM: Staphylococcus aureus
;; STRAIN: Clinical isolate SA-77
;; US-08-920-827-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 TTCTTATTAGACCTGTGTAGTACCTTAGAAACCAACACCACTAGA 54
1639 TTGCTGATGACGATTTTATTATTAATGATGATGCCAATATACCACTATA 1692

;; RESULT 8
;; US-08-921-177-4
;; Sequence 4, Application US/08921177
;; Patent No. 5798211
;; GENERAL INFORMATION:
;; APPLICANT: Ohno, Tsuneya
;; APPLICANT: Matsuhisa, Akio
;; APPLICANT: Uehara, Hirotsugu
;; APPLICANT: Eda, Soji
;; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/921,177
;; FILING DATE: 29-AUG-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/362,577
;; FILING DATE: 27-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rin-Laures, Li-Hsien
;; REGISTRATION NUMBER: 33,547
;; REFERENCE/DOCKET NUMBER: 19036/32420
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-921-177-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 1 TTCCTATTAGACCTGTGTAGTACCTTGAAACCAACACAACTTACA 54
DB 1639 TTGCTGAATGAGCATTTTATTTTAATATGATAGCCAAATATTAACAAGCTATA 1692

RESULT 9

US-08-362-577C-4
Sequence 4, Application US/08362577C
Patent No. 5807673

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-362-577C-4

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-362-577C-4

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-362-577C-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 1 TTCCTATTAGACCTGTGTAGTACCTTGAAACCAACACAACTTACA 54
DB 1639 TTGCTGAATGAGCATTTTATTTTAATATGATAGCCAAATATTAACAAGCTATA 1692

DB 1639 TTGCTGAATGAGCATTTTATTTTAATATGATAGCCAAATATTAACAAGCTATA 1692

RESULT 10

US-08-920-828-4
Sequence 4, Application US/08920828
Patent No. 5853998

GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-828-4

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-828-4

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-828-4

RESULT 11

US-08-755-587-4
Sequence 4, Application US/08755587
Patent No. 6045997

GENERAL INFORMATION:

APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the BRCA2 cancer
identification and sequencing of the BRCA2 cancer

TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-755-587-4

Query Match 34.6%; Score 21.8; DB 3; Length 96;
Best Local Similarity 61.4%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 4 CTTATTAAAGCCTGTGTAGTCTTGAAGAAACCAACACACAACTTAAAGATCC 60
Db 23 CTTATTAAAGATTTGACAGATATATAGAAATCAAGAAATCTTAAAGCTTC 79

RESULT 12
US-08-755-587-28
Sequence 28, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587

FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 483..576
FEATURE:
NAME/KEY: exon
LOCATION: 481..576
US-08-755-587-28

Query Match 34.6%; Score 21.8; DB 3; Length 1056;
Best Local Similarity 61.4%; Pred. No. 48;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 4 CTTATTAAAGCCTGTGTAGTCTTGAAGAAACCAACACAACTTAAAGATCC 60
Db 503 CTTATTAAAGATTTGACAGATATATAGAAATCAAGAAATCTTAAAGCTTC 559

RESULT 13
US-08-755-587-1
Sequence 1, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9

FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1917 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-755-587-1

Query Match 34.6%; Score 21.8; DB 3; Length 1917;
Best local Similarity 61.4%; Pred. No. 54;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4 TTATTTAAGCACTGTGTAGTACTTGAACCAACACACCACTAGAGATCC 60
DB 657 CTTATTAATGATTTTGACAGATATATGAATAATCAAGAAAATCTTAAGGCTTC 713

RESULT 14
US-08-840-466A-20
Sequence 20, Application US/08840466A
Patent No. 6261561

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

Street, 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/840,466A

FILING DATE: 18-Apr-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laural S.

REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995-0029-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-08-840-466A-20

Query Match 34.6%; Score 21.8; DB 4; Length 3131;
Best local Similarity 61.4%; Pred. No. 60;

Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 1 TTCTTATTTAAGCACTGTGTAGTACTTGAACCAACACCACTAGAGGA 57
DB 47 TTGGTATTTACATATATGAGGAAATACATTAGAAACGAACTATGAAATAGAGGA 103

RESULT 15
US-09-696-188B-20
Sequence 20, Application US/09696188B
Patent No. 6406885

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

Street, 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/696,188B

FILING DATE: 26-Oct-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/840,466

FILING DATE: 1997-04-18

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laural S.

REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995-0029-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-09-696-188B-20

Query Match 34.6%; Score 21.8; DB 4; Length 3131;
Best local Similarity 61.4%; Pred. No. 60;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 1 TTCTTATTTAAGCACTGTGTAGTACTTGAACCAACACCACTAGAGGA 57
DB 47 TTGGTATTTACATATATGAGGAAATACATTAGAAACGAACTATGAAATAGAGGA 103
Search completed: May 11, 2003, 03:07:40
Job time : 30.2632 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 62.3684 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-12
Perfect score: 63
Sequence: 1 ttccctatttaagcactgtg.....aacacactagagatccccc 63

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	100.0	63	US-09-963-803-12	Sequence 12, Appl
2	60	95.2	317	US-09-963-803-3	Sequence 3, Appl
3	60	95.2	371	US-09-963-803-5	Sequence 5, Appl
4	60	95.2	393	US-09-963-803-19	Sequence 19, Appl
5	60	95.2	462	US-09-963-803-20	Sequence 20, Appl
6	60	95.2	600	US-09-963-803-22	Sequence 22, Appl
7	49	77.8	243	US-09-963-803-1	Sequence 1, Appl
8	49	77.8	392	US-09-963-803-21	Sequence 21, Appl
9	46	73.0	301	US-09-963-803-7	Sequence 7, Appl
10	46	73.0	348	US-09-963-803-4	Sequence 4, Appl
11	46	73.0	398	US-09-963-803-6	Sequence 6, Appl
12	46	73.0	472	US-09-963-803-25	Sequence 25, Appl
13	46	73.0	541	US-09-963-803-24	Sequence 24, Appl
14	46	73.0	604	US-09-963-803-23	Sequence 23, Appl
15	25.2	40.0	373	US-09-796-692-6196	Sequence 6196, Ap
16	25.2	40.0	373	US-10-040-862-6196	Sequence 6196, Ap
17	24.8	39.4	815	US-09-925-300-302	Sequence 302, App
18	24.4	38.7	2915	US-10-028-072-209	Sequence 209, App
19	24.4	38.7	2915	US-10-121-049-209	Sequence 209, App

20	24.4	38.7	2915	US-10-123-904-209	Sequence 209, App
21	24.4	38.7	2915	US-10-140-470-209	Sequence 209, App
22	24.4	38.7	2915	US-10-175-746-209	Sequence 209, App
23	24.4	38.7	2915	US-10-176-918-209	Sequence 209, App
24	24.4	38.7	2915	US-10-176-921-209	Sequence 209, App
25	24.4	38.7	2915	US-10-137-865-209	Sequence 209, App
26	24.4	38.7	2915	US-10-140-474-209	Sequence 209, App
27	24.4	38.7	2915	US-10-142-431-209	Sequence 209, App
28	24.4	38.7	2915	US-10-143-114-209	Sequence 209, App
29	24.4	38.7	2915	US-10-140-002-209	Sequence 209, App
30	24.4	38.7	2915	US-10-142-419-209	Sequence 209, App
31	24.4	38.7	2915	US-10-123-262-209	Sequence 209, App
32	24.4	38.7	2915	US-10-142-423-209	Sequence 209, App
33	24.4	38.7	2915	US-10-121-050-209	Sequence 209, App
34	24.4	38.7	2915	US-10-141-755-209	Sequence 209, App
35	24.4	38.7	2915	US-10-143-032-209	Sequence 209, App
36	24.4	38.7	2915	US-10-123-108-209	Sequence 209, App
37	24.4	38.7	2915	US-10-123-336-209	Sequence 209, App
38	24.4	38.7	2915	US-10-123-261-209	Sequence 209, App
39	24.4	38.7	2915	US-10-140-921-209	Sequence 209, App
40	24.4	38.7	2915	US-10-140-928-209	Sequence 209, App
41	24.4	38.7	2915	US-10-121-045-209	Sequence 209, App
42	24.4	38.7	2915	US-10-123-292-209	Sequence 209, App
43	24.4	38.7	2915	US-10-123-903-209	Sequence 209, App
44	24.4	38.7	2915	US-10-124-819-209	Sequence 209, App
45	24.4	38.7	2915	US-10-124-822-209	Sequence 209, App

ALIGNMENTS

RESULT 1
US-09-963-803-12
Sequence 12, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 18432042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 63
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: S5
US-09-963-803-12
Query Match 100.0%; Score 63; DB 9; Length 63;
Best Local Similarity 100.0%; Pred. No. 8.4e-14;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Query 1 TTCTTATTTAAGACTGTGTAGTACTTGAAGAACACACACACACCTAGAGGATCC 60
Db 1 TTCTTATTTAAGACTGTGTAGTACTTGAAGAACACACACACACCTAGAGATCC 60
QY 61 CCG 63
Db 61 CCG 63
RESULT 2
US-09-963-803-3
Sequence 3, Application US/09963803

```
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIORITY FILING DATE: 2001-09-26
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 2000-10-05
PRIORITY FILING DATE: 2000-10-05
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116
NAME/KEY: promoter
LOCATION: (1)..(317)
OTHER INFORMATION:
US-09-963-803-3

Query Match          95.2%; Score 60; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAAACCAACAACAACCTAGAGATCC 60
DB 258 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAAACCAACAACAACCTAGAGATCC 317

RESULT 3
US-09-963-803-5
Sequence 5, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIORITY FILING DATE: 2001-09-26
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 2000-10-05
PRIORITY FILING DATE: 2000-10-05
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 371
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1146
NAME/KEY: promoter
LOCATION: (1)..(371)
OTHER INFORMATION:
US-09-963-803-5

Query Match          95.2%; Score 60; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 4
US-09-963-803-19
Sequence 19, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIORITY FILING DATE: 2001-09-26
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 2000-10-05
PRIORITY FILING DATE: 2000-10-05
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1162
NAME/KEY: promoter
LOCATION: (1)..(393)
OTHER INFORMATION:
US-09-963-803-19

Query Match          95.2%; Score 60; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAAACCAACAACAACCTAGAGATCC 60
DB 334 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAAACCAACAACAACCTAGAGATCC 393

RESULT 5
US-09-963-803-20
Sequence 20, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIORITY FILING DATE: 2001-09-26
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 1999-03-29
PRIORITY FILING DATE: 2000-10-05
PRIORITY FILING DATE: 2000-10-05
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1163
NAME/KEY: promoter
LOCATION: (1)..(462)
OTHER INFORMATION:
US-09-963-803-20

Query Match          95.2%; Score 60; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TTCCTATTAAAGCACTGTGTAGTAGCTTAGAAAACCAACAACAACCTAGAGATCC 60
|||||
```

Db 403 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 462

RESULT 6

US-09-963-803-22
; Sequence 22, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1165
; NAME/KEY: promoter
; LOCATION: (1)..(600)
; OTHER INFORMATION:
US-09-963-803-22

Query Match 95.2%; Score 60; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 60
Db 541 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 600

RESULT 7

US-09-963-803-1
; Sequence 1, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 243 bp Fragment from the intergenic region of commelina yellow
; OTHER INFORMATION: title virus
; NAME/KEY: promoter
; LOCATION: (1)..(243)
; OTHER INFORMATION:
US-09-963-803-1

Query Match 77.8%; Score 49; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 49
Db 195 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 243

RESULT 8

US-09-963-803-21
; Sequence 21, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1164
; NAME/KEY: promoter
; LOCATION: (1)..(392)
; OTHER INFORMATION:
US-09-963-803-21

Query Match 77.8%; Score 49; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 49
Db 334 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACCACTTAGAGATCC 382

RESULT 9

US-09-963-803-7
; Sequence 7, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1154
; NAME/KEY: promoter
; LOCATION: (1)..(301)
; OTHER INFORMATION:
US-09-963-803-7

Query Match 73.0%; Score 46; DB 9; Length 301;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
DB 256 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 301

RESULT 10
US-09-963-803-4
Sequence 4, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 348
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1117
NAME/KEY: promoter
LOCATION: (1)..(348)
OTHER INFORMATION:
US-09-963-803-4

Query Match 73.0%; Score 46; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
DB 303 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 348

RESULT 11
US-09-963-803-6
Sequence 6, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 398
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1147
NAME/KEY: promoter

LOCATION: (1)..(398)
OTHER INFORMATION:
US-09-963-803-6

Query Match 73.0%; Score 46; DB 9; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
DB 353 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 398

RESULT 12
US-09-963-803-25
Sequence 25, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25
LENGTH: 472
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1169
NAME/KEY: promoter
LOCATION: (1)..(472)
OTHER INFORMATION:
US-09-963-803-25

Query Match 73.0%; Score 46; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
DB 427 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 472

RESULT 13
US-09-963-803-24
Sequence 24, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 541
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:


```

; OTHER INFORMATION: promoter MPR1168
; FEATURE:
; NAME/KEY: Promoter
; LOCATION: (1)..(541)
; OTHER INFORMATION:
US-09-963-803-24

```

```

Query Match          73.0%; Score 46; DB 9; Length 541;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 15 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 60
DB 496 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 541

```

```

RESULT 14
US-09-963-803-23
; Sequence 23, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow m
; TITLE OF INVENTION: virus and casava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: PR 98/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MPR1167
; NAME/KEY: promoter
; LOCATION: (1)..(604)
; OTHER INFORMATION:
US-09-963-803-23

```

```

Query Match          73.0%; Score 46; DB 9; Length 604;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 15 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 60
DB 559 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 604

```

```

RESULT 15
US-09-796-692-6196
; Sequence 6196, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algaier, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27

```

```

; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6196
; LENGTH: 373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (316)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6196

```

```

Query Match          40.0%; Score 25.2; DB 9; Length 373;
Best Local Similarity 65.5%; Pred. No. 9.5;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 5 TTATTAGACCTGTGTAGTACTTGAACCAACCAACCACTAGAGATC 59
DB 289 TTCTTATTCATCTCTGTGAAGCTTANAAACCAACCAAGCTGTGTGCTTC 343

```

Search completed: May 11, 2003, 06:11:42
Job time : 63.3684 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1012.74 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttcctattactagcactgtf.....aacacactagagatcccccg 63

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	30	47.6	901 12	BG294121 602390966
2	29.4	46.7	679 9	AL777174 AL777174
3	29	46.0	526 17	AG025881 Homo sapi
4	27.4	43.5	472 14	BQ968546 QH34E11
5	27.2	43.2	546 17	AZ652445 IM0525J11
6	27	42.9	664 13	BM575510 170006871

C	7	27	42.9	724 13	BM617342
C	8	26.6	42.2	514 9	AA053822
C	9	26.6	42.2	549 13	BM094741 sa53f01.r
C	10	26.6	42.2	843 14	BM0963329
C	11	26.4	41.9	343 12	BG461281
C	12	26.4	41.9	501 17	AZ407729
C	13	26.2	41.6	389 9	AA757510
C	14	26.2	41.6	750 13	BI917143
C	15	26.2	41.6	762 14	B0745595
C	16	25.8	41.0	769 17	BH714488
C	17	25.6	40.6	458 17	AG685997
C	18	25.6	40.6	509 12	BF156127
C	19	25.6	40.6	540 17	AQ080438
C	20	25.6	40.6	549 13	BM037729
C	21	25.6	40.6	559 13	BM155751
C	22	25.6	40.6	662 14	B0616770
C	23	25.6	40.6	677 13	BI903774
C	24	25.6	40.6	708 17	AQ098421
C	25	25.6	40.6	716 10	AV705148
C	26	25.4	40.3	492 17	AZ312818
C	27	25.4	40.3	638 17	AZ248833
C	28	25.4	40.3	858 12	BG617584
C	29	25.4	40.3	911 12	BF794977
C	30	25.2	40.0	380 14	BQ185082
C	31	25.2	40.0	771 17	CNS05RVV
C	32	25	39.7	459 10	AM085143
C	33	25	39.7	534 17	BH823752
C	34	25	39.7	601 17	BH828901
C	35	25	39.7	817 12	BE884936
C	36	25	39.7	931 17	AZ539154
C	37	25	39.7	1220 14	BQ396668
C	38	24.8	39.4	264 12	BQ67453
C	39	24.8	39.4	286 14	R34801
C	40	24.8	39.4	302 12	BG985557
C	41	24.8	39.4	362 14	BQ488731
C	42	24.8	39.4	367 14	BM784926
C	43	24.8	39.4	378 12	BF079516
C	44	24.8	39.4	440 10	AW781100
C	45	24.8	39.4	457 17	AQ225617

ALIGNMENTS

RESULT 1
LOCUS BG294121/c 901 bp mRNA linear EST 21-FEB-2001
DEFINITION 602390966F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502925 5',
ACCESSION BG294121
VERSION BG294121.1 GI:13054439
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepo Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LLMA0372 row: k column: 22
High quality sequence stop: 718.
Location/Qualifiers

FEATURES
source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4502925"
 /clone_1ib="NIH_MGC_94"
 /libuse_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 234 a 195 c 204 g 268 t

Query Match 47.6%; Score 30; DB 12; Length 901;
 Best Local Similarity 67.7%; Pred. No. 9.3;

Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 TTCCTATTAGACCTGTGTAGTACCTTAGAAAACCAACACACACCTAGAGATCC 60

Db 266 TTCCTATTAGACACGAGTGTGTGTGTTGAAAACCAACCAACCTTAGAAGACA 207

Qy 61 CC 62

Db 206 CC 205

RESULT 2

AL777174/c AL777174 679 bp mRNA linear EST 25-JUN-2002

LOCUS AL777174 XGC-gaetrula silurana tropicalis cDNA clone TGa8081f17 5,

DEFINITION mRNA sequence.

ACCESSION AL777174 GI:21562878

VERSION EST.

KEYWORDS western clawed frog.

SOURCE Silurana tropicalis

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae;

Xenopodinae; Silurana.

REFERENCE 1 (bases 1 to 679)

AUTHORS Taylor,R., Ashurst,J.L., Cronling,M.D.R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2002

JOURNAL Unpublished (2001)

COMMENT Contact: Taylor R

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: tropesanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS_SEQUENCE_ID: TGa8081f17.plksp6

Sequencing primer: PlkSP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1. 679

/organism="Silurana tropicalis"

/db_xref="taxon:8364"

/clone="TGa8081f17"

/clone_1ib="XGC-gaetrula"

/dev_stage="gaetrula (stages 10.5-13 mixed)"

/lab_host="Escherichia coli XL1-blue"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from 5' end of poly A+ RNA from stages

10-13 gastrulae. EcoRI/NotI cut cDNA was then ligated

into pCS107 with EcoRI at the 5' end and NotI at the 3'

end."

BASE COUNT 213 a 134 c 116 g 216 t

ORIGIN

Query Match 46.7%; Score 29.4; DB 9; Length 679;

Best Local Similarity 70.9%; Pred. No. 15;

Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 4 CTTATTAGACCTGTGTAGTACCTTAGAAAACCAACACACACCTAGAGAT 58

Db 648 CTCATTGAAACCTGTGTGTGTTGAAAACCAACCAACCTTAGAAGAT 594

RESULT 3

AG025881 526 bp DNA linear GSS 21-DEC-1999

LOCUS AG025881 Homo sapiens DNA, chromosome 8q23, reverse end of BAC clone:

DEFINITION KB1562D12, genomic survey sequence.

ACCESSION AG025881 GI:6624572

VERSION GSS.

KEYWORDS Homo sapiens DNA, clone_1ib:KEIO BAC library clone:KB1562D12.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 526)

AUTHORS Shimizu,N. and Asakawa,S.

TITLE The BAC end sequence of Homo sapiens BAC clone KB1562D12

JOURNAL Published only in Database (1999)

REFERENCE 2 (bases 1 to 526)

AUTHORS Shimizu,N. and Asakawa,S.

TITLE Direct Submission

JOURNAL Submitted (15-DEC-1999) Nobuyoshi Shimizu, Keio University, School

of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo

160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp.

Tel:81-3-3351-2370, Fax:81-3-3351-2370)

Location/Qualifiers

1. 526

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="8q23"

/clone="KB1562D12"

/clone_1ib="KEIO BAC library"

/note="This sequence is reverse end of BAC clone

KB1562D12."

BASE COUNT 112 a 146 c 90 g 178 t

ORIGIN

Query Match 46.0%; Score 29; DB 17; Length 526;

Best Local Similarity 71.7%; Pred. No. 20;

Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 8 TTTAGACCTGTGTAGTACCTTAGAAAACCAACACACACCTAGAGATCC 60

Db 343 TTTAAACCTTCAGTGTGTGCTGCTGAGATTTACACACACCTTAGAATCC 395

RESULT 4

BO968546/c BO968546 472 bp mRNA linear EST 21-AUG-2002

LOCUS QHB34E11.yg.ab1 OH ABCDI sunflower RHA801 Helianthus annuus cDNA

DEFINITION clone QHB34E11, mRNA sequence.

ACCESSION BO968546

VERSION BO968546.1 GI:22386067

KEYWORDS EST.

SOURCE common sunflower.

ORGANISM Helianthus annuus

REFERENCE Helianthus annuus

1 (bases 1 to 472)

AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,

Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison

,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,

Church,S., Jackson,L. and Bradford,K.

Lettuce and Sunflower ESTs from the Compositeae Genome Project

http://compgenomics.ucdavis.edu/

Unpublished (2002)

COMMENT Contact: Alexander Kozik [R.W.Michelmore]

Email: akozik@atgc.org [michel.moreovegma1.ucdavis.edu]
 belongs to config QH_CA_Config1526, see <http://cgrpd.ucdavis.edu/>
 for details.

Location/Qualifiers
1. .472

Query Match	43.5%	Score 27.4:	DB 14,	Length 472;
Best Local Similarity	68.5%	Pred. No. 67;		
Matches 37, Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

Qy 1 TTCCTTATTTAAGCAGTTGTGTAAGCTTGAAAAACCAACAACAACCTAGA 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 TTACATCATTAACACCCTTTGAAGTAGCATAGTAACCNACATATGAAGAGAGA 24

LOCUS	AZ652445	546 bp	DNA	linear	14-DEC-2000
DEFINITION	1M0525u11R Mouse 10kb plasmid UUGC1M library				
DESCRIPTION	clone UUGC1M0525u11 R, DNA sequence.				

AZ652445.1 GI:11788967
GSS.

ORGANISM
Mus musculus
+house mouse.
SOURCE
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (house) 1 (no. Euk)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,

Mouse whole genome scaffolding with paired end reads from 10kbp plasmid inserts

Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

```
InsertLength: 10000    Std Error: 0.00
Plate: 0525    row: J    column: 11
Seq primer: CACACAGCAAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 546.
```

Location/Qualifiers

```
1. 546
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UNC1M0525j11"
/clone_lib="mouse 10kb plasmid unc1m library"
sex="Male"
```

1A2 host-E. coli strain XL10-Gold, 11-resistant, *E. coli* Vector: PMD42-*lacZ*. Purified genomic DNA from *M. musculus* C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrolytically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-treated with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gels. Electrophoresis. Vector DNA was prepared from a derivative of pMD42 [gill473111(gplA123072.1)], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match	43.2%	Score	27.2;	DB	17;	Length	546;
Best Local Similarity	72.9%	Pred. No.	78;				
Matches	35;	Conservative	0;	Mismatches	13;	Indels	0;
						Gaps	0;

Qy 1 TTCCTATTTAAGCACC TTGTGTAGTACCTTAGAAAAACCAACAACAA 48
| | | | | | | | | | | | | | | | | | | | | |
Db 217 TACCATATTTGAGAATTGTTTAATCACTTAGAAAAACAACCCAA 170

LOCUS	664 bp	mRNA	linear	EST 22-FEB-2002
DEFINITION	BMS75510	170006871198 A.Gam.ad.cDNA	blood1	Anopheles gambiae
DEFINITION	19600449636818	5', mRNA sequence.		

75510.1 GI:18863977

SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

(bases 1 to 664)
t,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab

Published (2002)
Contact: Holt R.A.

W. Gude Dr., Rockville, MD 20850, USA
: 2404533151

```

: 2404534580
il: HoltRA@celera.com
te: NU01004AAV . row: D column: 20
primer: M13 Reverse.

```

Location/Qualifiers
1. .664

```

/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std
Chromosome)"
/db_xref="caxon:7165"
/clone="19600449696818"
/clone_lib="A. Gam. ad. cDNA: blood1"

```

```

/dev stage="Adult"
/lab host="DH10b"
/note="Vector: pSport1, Site_1: SalI, Site_2: NotI, Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. CDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mri.org)"
182 a 155 c 143 g 184 t

```

Query Match	42.9%	Score 27	DB 13	Length 664
Best Local Similarity	70.6%	Pred. No. 91		
Matches	36	Conservative	0	Mismatches 15, Indels 0, Gaps 0,
7	ATTATAGACCTGTGTAGTACCTTGTGAAACACACACACACCTTAGACA	57		
DB	526	ATTATAGCTCTGTGTGACGTACCAATATATATACCAACACACACCTTAGACA	476	

RESULT 7	BM617342/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	BM617342	17000687151386 A.Gam.ad.cDNA.bloodi	724 bp mRNA linear EST 25-FEB-2002					
		1960044971920 5, mRNA sequence.						
	BM617342							
	BM617342.1	GI:18915564						
	EST.							
	African malaria mosquito.							
	Anopheles gambiae							
	Anopheles gambiae							

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 724)	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlabach, R., Colling, F.H., Venter, J.C. and Hoffman, S.L.	Ceiera Anopheles gambiae ESR project	Unpublished (2002)	Contact: Holt R.A.

45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: Holt@aeclera.com
Place: NU010049UR row: K column: 10
Seq primer: MJ Reverse
Location/Qualifier

```

Location/Qualifiers
1. 724
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449731920"
/clone_1ib="A.Gam.ad.CDNA.blood1"
/dev_stage="Adult"
/lab_host="Dh10b"
/note="Vector: pSport1; Site 1: Sal1; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 2p
hours after human blood feeding. CDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mri.org)"
BASE COUNT 216 a 154 c 168 g 186 t
ORIGIN

```

Query Match	42.9%	Score 27	DB 13	Length 724
Best Local Similarity	70.6%	Pred. No. 92		
Matches	36	Conservative	0	Mismatches 15; Indels 0; Gaps 0
Oy	7	ATTTAGACCTTGTGTAGTACGTAGAAAAACCAACCAACACCTTAGAGA	57	
Db	92	ATTGAGCTTGTGTGAGTAAACAATAATTAACAAACCAACCTTAGAGA	42	

RESULT	8
AA053822	
LOCUS	514 bp mRNA linear EST 02-FEB-1997
DEFINITION	zef3f01.r1 Soares retina N2bHR Homo sapiens CDNA clone
ACCESSION	AA053822
VERSION	AA053822.1 GI:1544757
KEYWORDS	EST.
SOURCE	human.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (Pages 510 to 514)	Becker, M., Bonaldo, M.F., Chiappelli, B., Hillier, L., Lennon, G., Chissee, S., Dietrich, N., Dubquet, T., Favello, A., Gish, N., Hawkins, M., Hultman, M., Kucaba, T., Laczy, M., Le, N., Merdis, R., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaatis, Underwood, K., Wohlmann, P., Watson, R., Wilson, R. and Marra, M.	Generation and analysis of 280,000 human expressed sequence tags	Genome Res. 6 (3), 807-828 (1996)	Contact: Wilson RK 97044478

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Putative full length read
The vector to vector length is 1104
Insert Length: 2734 Std Error: 0.00
Seq primer: -26ml rev2 from Amersham
High quality sequence stop: 452.

FEATURES
SOURCE

```

/organism="Homo sapiens"
/db_xref="GDB:128892"
/db_xref="taxon:9606"
/clone="IMAGE:38065"
/clone_id="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pTR73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACCAATCTGAAATGGAGGAGCGCCGCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTR73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."

```

	Query Match	Similarity	Score	No.6:	DB 9:	Length
	Best Local	Similarity	66.7%	Pred.	No.1,2e+02:	
	Matches	38;	Conservative	0;	Mismatches	19; Indels 0; Gaps 0;
OY	ATTTAAGCACTTGTTACTAGCTTAGAAAAACCAACAACAACCTTGAAGATCCCCG	63				
Dd	ATTATATGCCCTGTGTGTAGAAAGCTAGAAAAAAAAAAGAAGAAAAGAAATCTCTAG	403				

Db 347 ATTATGCTCTGTGTGAAGCTAGAAAAAAAAAGAAAGAACTCTAG 403

BM094741
LOCUS 549 bp mRNA linear EST 30-NOV-2001
DEFINITION B094741.1 Y1-Gm-cl066 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl066-3301 5' similar to TR:Q9SH35 Q9SH35 F2K11.13. ; mRNA
sequence.
ACCESSION BM094741.1 GI:17023707
VERSION BM094741.1
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 549)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
JOURNAL Public Soybean EST Project
COMMENT Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Possible reversed clone: similarity on wrong strand This clone is
available through: Resgen, Invitrogen Corp. 2130 South Memorial
Parkway Huntsville, AL 35801 For further information call: (800
)-533-4133 or contact via email: ccr@resgen.com
High quality sequence stop: 421.
Location/Qualifiers
1. 549
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl066-3301"
/clone_1lb="Gm-cl066"
/tissue_type="Leaf and shoot tip, salt stressed, 2 week
old seedling"
/lab_host="DH10B"
/note="Vector: Bluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from unexpanded leaves and the shoot tips of 2 week old
seedlings from the cultivar Williams. The 2 week old
seedlings were salt stressed in a solution of 500mM NaCl
for 3 days prior to harvesting. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the Bluescript vector. The ligated cDNA fragments were
transformed into DH10B host cells (GibcoBRL). This library
was constructed in the laboratory of Dr. Randy
Shoemaker."

BASE COUNT 187 a 124 c 82 g 156 t
ORIGIN

Query Match 42.2%; Score 26.6; DB 13; Length 549;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 TTCCTATTAGACCTGTGTAGTACCTAGAAACCAACACACACAC 49
Db 9 TTTTATTTTAAAGCAATGTCTGTACATTAATTACAGACACAAAC 57

RESULT 10
BQ963329/c

BQ963329
LOCUS 843 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT 10050135 NIH MGC 134 Mus musculus cDNA clone
IMAGE:6508493 5', mRNA sequence.
ACCESSION BQ963329
VERSION BQ963329.1 GI:22378807
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 843)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM14074 row: e column: 06
High quality sequence stop: 467.
Location/Qualifiers
1. 843
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6508493"
/clone_1lb="NIH MGC 134"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-Sport1.1(ccdb); Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dT. Average
insert size 1.7 kb. Constructed by Resgen, Invitrogen
Corp. Note: this is a NIH MGC Library."

BASE COUNT 214 a 204 c 191 g 233 t
ORIGIN

Query Match 42.2%; Score 26.6; DB 14; Length 843;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 6 TATTATGACCTGTGTAGTACCTAGAAACCAACACACACACTGAGGATCCCC 62
Db 142 TATGTGTGACGTTGTCTATACACACAGAAACCAACCAACACACAGTGTATCCC 86

RESULT 11
BQ461281
LOCUS 343 bp mRNA linear EST 21-APR-2001
DEFINITION RST44060 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ461281
VERSION BQ461281.1 GI:13749787
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith
J., Danzig, J., and Ducar, M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL 21227151
MEDLINE
CONTACT: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900
Fax: 216 361 9596
Email: ecat@etherys.com
Location/Qualifiers

FEATURES

SOURCE

1. 343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 135 a 81 c 46 g 81 t
ORIGIN

Query Match 41.9%; Score 26.4; DB 12; Length 343;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 7 ATTAAAGCACTGTGTAGTCTTGAAGAACCAACCAACCACTAGAGAT 58
Db 4 ATCAAGAGCTGTATATATATTGGAAACCAACCAACCAACGAGAGAT 55

RESULT 12 501 bp DNA linear GSS 03-OCT-2000
LOCUS AA757510/c
DEFINITION I00178002R Mouse 10kb plasmid UGCM1 library Mus musculus genomic
ACCESSION I00178002R
VERSION AA757510.1 GI:10531742
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE 1 (bases 1 to 501)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT

Unpublished (2000)
Contact: Robert B. Weis
University of Utah Genome Center
Rm. 308, Biomedical polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0178 row: 0 column: 02
Seq primer: CACACAGAAACACACTGTGACC
Class: plasmid ends
High quality sequence stop: 501.

FEATURES

SOURCE

1. 501
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCM178002"
/clone_lib="Mouse 10kb plasmid UGCM1 library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male); was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|47321149|AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 116 a 112 c 125 g 148 t
ORIGIN

Query Match 41.9%; Score 26.4; DB 17; Length 501;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 7 ATTAAAGCACTGTGTAGTCTTGAAGAACCAACCAACCACTAGAGAT 58
Db 422 ATTAAAGCACTGTGTCTTGTGGAAAGCAAGAAACCAAGCTTAGACTT 473

RESULT 13 389 bp mRNA linear EST 23-JAN-1998
LOCUS AA757510
DEFINITION Z937809.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone
IMAGE:395512 3', mRNA sequence.
ACCESSION AA757510
VERSION AA757510.1 GI:2805373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 389)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Kitzman, D., Kucab, T., Lacy, M., Le, N., Lennon, G., Mair, M., Martin, J., Moore, B., Schellberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterson, R. and Wilson, R.

TITLE Washu-NCI human EST Project

JOURNAL

COMMENT

Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 353.

FEATURES

SOURCE

1. 389
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:1301414"
/db_xref="taxon:9606"
/clone="IMAGE:395512"
/clone_lib="Soares pineal gland N3HPG"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pRTT3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'] double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT	107 a	79 c	69 g	134 t
ORIGIN				
Query Match	41.6%; Score 26.2; DB 9; Length 389;			
Best Local Similarity	67.3%; Pred. No. 1.7e+02;			
Matches 37; Conservative	0; Mismatches 18; Indels 0; Gaps 0;			
Qy	7 ATTATGACCTGCTGTAGTCTTGAAGAAACCAACACCAACCTAGAGATCC 61			
Db	77 ATTATGCTCTGTGTAGAGAGCTAGAAAAAAGAGAAAAAAGAGAACTCTC 23			
RESULT 14				
LOCUS	BI917143 750 bp mRNA linear EST 16-OCT-2001			
DEFINITION	603181563p1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5245646 5',			
ACCESSION	BI917143			
VERSION	BI917143.1 GI:16181105			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 750)			
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cga@bbs-rcmail.nih.gov			
	Tissue Procurement: Life Technologies, Inc.			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)			
	DNA Sequencing by: Incyte Genomics, Inc.			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LNLN at:			
	http://image.lnl.gov			
	Plate: LLM11619 row: n column: 15			
	High quality sequence stop: 748.			
FEATURES				
source	location/Qualifiers			
	1..750			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:5245646"			
	/clone_id="NIH_MGC_121"			
	/lab_host="DH10B"			
	/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;			
	Site 2: EcoRV (destroyed); RNA source anonymous pool of 3			
	fetal brains, female age 20 weeks, female age 24 weeks,			
	and male age 26 weeks. Library is oligo-dT primed and			
	directionally cloned (EcoRV site is destroyed upon			
	cloning). Average insert size 1.7 kb, insert size range			
	0.7-3.5 kb. Library is normalized and enriched for			
	full-length clones and was constructed by C. Gruber			
	(Invitrogen). Research Genetics tracking code 017. Note:			
	this is a NIH-MGC Library."			
BASE COUNT	225 a	139 c	170 g	216 t
ORIGIN				
Query Match	41.6%; Score 26.2; DB 13; Length 750;			
Best Local Similarity	67.3%; Pred. No. 1.7e+02;			
Matches 37; Conservative	0; Mismatches 18; Indels 0; Gaps 0;			
Qy	5 TTATTTAAGCACTGTGTAGTCTTGAAGAAACCAACCAACACTAGAGATC 59			
Db	431 TTCTTATTCATCTCTGTGAAGCTTAGAAACCAACCAACACTGTGTGTTTC 485			
RESULT 15				
LOCUS	BO745595 762 bp mRNA linear EST 17-JUL-2002			
DEFINITION	UI-M-EMO-bxb-g-17-0-UI.r1 NIH_MGC Mus musculus cDNA clone			
	IMAGE:5708200 5', mRNA sequence.			

```

ACCESSION      BO745595
VERSION        BO745595.1
KEYWORDS       GI:21892382
SOURCE         EST.
ORGANISM       house mouse.
               Mus musculus.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE      1 (bases 1 to 762)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgaps-remail.nih.gov
               Tissue Procurement: Dr. James Lin, University of Iowa
               cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
               cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNL at:
               http://image.llnl.gov
               This clone was contributed by the Brain Molecular Anatomy Project
               (BMAP)

FEATURES
source         Location/Qualifiers
               1..762
               /organism="Mus musculus"
               /strain="C57BL/6"
               /db_xref="taxon:10090"
               /clone="IMAGE:5708200"
               /clone_1lb="NIH BMAP EM0"
               /tissue_type="Whole Brain"
               /dev_stage="embryo 15.5 dpc"
               /lab_host="DH10B (T1 phage resistant)"
               /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
               Site 2: Not I; The library was constructed according to
               Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
               1996. Denatured mRNA was size fractionated on a 1% agarose
               gel. First strand cDNA synthesis was primed with an
               oligo-dT primer containing a Not I site. Double stranded
               cDNA was size selected according to mRNA size fraction,
               ligated with EcoR I adaptor, digested with Not I, and then
               cloned directionally into pYX-Asc vector. The library tag
               sequence located between the Not I site and the polyA tail
               , is GGGCGTGA. This library was created for the
               University of Iowa Mouse Brain Molecular Anatomy Project
               (BMAP): 'Gene Discovery in the Developing Mouse Nervous
               System', supported by National Institutes of Mental Health
               (NIMH), Hemin Chhn, Ph.D., program coordinator."

BASE COUNT     171 a      148 c      198 g      244 t      1 others
ORIGIN
Query Match    41.6%; Score 26.2; DB 14; Length 762;
Best Local Similarity 67.3%; Pred. No. 1.7e+02;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY             9 TTAAGCACTTGCTAGTAGCTTAGAAAAACAACAACAACCTTAGAGAGATCCCG 63
               |||||
Db             490 TCAAGCACTTATTAATAATTCACAAACACCAACAACCATCATGCACG 436
               |||||

Search completed: May 11, 2003, 04:54:18
Job time : 1017.74 secs

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 153.644 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgcgcagcactgcatgatt.....aaggatgacatgcacact 79

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	79	21	AAA96848
2	79	100.0	79	21	AAA96472
3	75	94.9	259	21	AAA96477
4	75	94.9	259	21	AAA96465
5	70	88.6	472	21	AAA96859
6	70	88.6	541	21	AAA96858
7	69.6	88.1	393	21	AAA96853
8	69.6	88.1	462	21	AAA96854
9	69.6	88.1	600	21	AAA96856

C	10	69	87.3	392	21	AAA96855	Nucleotide sequenc
	11	68.2	86.3	604	21	AAA96857	Nucleotide sequenc
	12	68	86.1	371	21	AAA96839	Nucleotide sequenc
C	13	58	73.4	80	22	AA003420	CanV as-2/as-1 PCR
	14	58	73.4	226	22	AA003396	Mpr1133 promoter D
	15	58	73.4	229	22	AA003397	Mpr1134 promoter D
	16	58	73.4	332	22	AA003398	Mpr1135 promoter D
	17	58	73.4	472	22	AA003394	Mpr1130 promoter D
	18	55	69.6	80	22	AA003421	CanV as-2/as-1 PCR
	19	46	58.2	63	21	AAA96847	Directional desoxy
	20	46	58.2	63	21	AAA96471	Nucleotide sequenc
	21	46	58.2	280	21	AAA96463	Nucleotide sequenc
	22	46	58.2	301	21	AAA96841	Nucleotide sequenc
	23	46	58.2	398	21	AAA96840	Nucleotide sequenc
C	24	39.2	49.6	63	22	AA003422	CanV as-2/as-1 PCR
	25	39.2	49.6	219	22	AA003399	Mpr1136 promoter D
	26	39.2	49.6	282	22	AA003400	Mpr1137 promoter D
	27	39.2	49.6	315	22	AA003401	Mpr1138 promoter D
	28	39.2	49.6	381	22	AA003406	Mpr1200 promoter D
	29	39.2	49.6	455	22	AA003395	Mpr1131 promoter D
	30	39.2	49.6	505	22	AA003402	Mpr1139 promoter D
	31	39	49.4	189	22	AA010066	Amplified product
	32	39	49.4	240	21	AAA71895	Soybean RRS gene N
	33	39	49.4	309	21	AA087192	CanV35S promoter e
	34	39	49.4	309	24	AA094442	Caulliflower mosaic
	35	39	49.4	333	22	AA083764	Plasmid pMON295 Ca
	36	39	49.4	334	22	AA083769	Nucleotide sequenc
	37	39	49.4	338	20	AA099492	Caulliflower mosaic
	38	39	49.4	350	13	AA023533	CanV35S promoter (
	39	39	49.4	352	13	AA025080	CanV35S promoter,
	40	39	49.4	355	15	AA072688	Caulliflower mosaic
	41	39	49.4	356	17	AA032299	Caulliflower mosaic
	42	39	49.4	366	22	AA028368	Nucleotide sequenc
	43	39	49.4	400	20	AA099491	Sequence containin
	44	39	49.4	444	20	AA084666	35S CMV promoter s
	45	39	49.4	444	22	AA089658	Tomato spotted wil

ALIGNMENTS

RESULT 1	
AAA96848	
ID	AAA96848 standard; DNA; 79 BP.
XX	
AC	AAA96848;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Directional desoxynucleotide building block S7.
XX	
KW	Promoter; intergenic region; Comelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; ss.
XX	
OS	Synthetic.
XX	
PN	WO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
PF	29-MAR-2000; 2000WO-IB00370.
XX	
PR	29-MAR-1999; 99FR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
PI	Rance I, Gruber V, Theisen M;
XX	
DR	WPI; 2000-647238/62.
XX	
PT	Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 26; 91pp; English.
CC The present sequence represents a directional deoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comelina yellow mottle virus, and the second plant promoter originates
CC from the Casarava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 79 BP; 22 A; 15 C; 20 G; 22 T; 0 other;
Query Match 100.0%; Score 79; DB 21; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATGATGATATCAAGATTGATGATCTCCACTGACGTA 60
DB 1 CATGCTGAGACTAGTATGATGATATCAAGATTGATGATCTCCACTGACGTA 60
QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79
RESULT 2
ID AAA96472 standard; DNA; 79 BP.
XX
AC AAA96472;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the directional building block S7.
XX
KM petE promoter; chimeric promoter; transgenic plant; MPr1108;
XX plastocyanin gene promoter; PCR primer; ss.
XX
OS Synthetic.
XX
OS WO200056906-A1.
XX
PN 28-SEP-2000.
XX
PD 20-MAR-2000; 2000WO-IB00317.
XX
PF 22-MAR-1999; 99FR-0003635.
XX
PR (MERT-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI WPI; 2000-587667/55.
XX
DR WPI; 2000-587667/55.
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
XX
PS Claim 24; Page 76; 83pp; English.
XX
CC The present sequence represents the directional building block S7,
CC which is used to construct chimeric promoters of the invention in
CC PCR reactions. The specification describes a chimeric expression
CC promoter comprising a petE promoter of the pea plastocyanin gene, or

CC comprising a G box operably or functionally linked upstream of a
CC CAAT box, TATA box and transcription initiation site. The chimeric
CC promoters are used in expression vectors for producing transgenic
CC plants, such as dicotyledonous species, e.g. potato, tobacco, cotton,
CC lettuce, tomato, melon, cucumber, pea, rape, beetroot, or sunflower,
CC and monocotyledonous species, e.g. wheat, barley, oat, rice, or corn.
XX
SQ Sequence 79 BP; 22 A; 15 C; 20 G; 22 T; 0 other;
Query Match 100.0%; Score 79; DB 21; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATGATGATATCAAGATTGATGATCTCCACTGACGTA 60
DB 1 CATGCTGAGACTAGTATGATGATATCAAGATTGATGATCTCCACTGACGTA 60
QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79
RESULT 3
ID AAA96477 standard; DNA; 259 BP.
XX
AC AAA96477;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the promoter MPr1112.
XX
KM petE promoter; chimeric promoter; transgenic plant; MPr1112;
XX plastocyanin gene promoter; ss.
XX
OS Pisum sativum.
XX
OS WO200056906-A1.
XX
PN 28-SEP-2000.
XX
PD 20-MAR-2000; 2000WO-IB00317.
XX
PF 22-MAR-1999; 99FR-0003635.
XX
PR (MERT-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI WPI; 2000-587667/55.
XX
DR WPI; 2000-587667/55.
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
XX
PS Claim 25; Page 79; 83pp; English.
XX
CC The present sequence represents the chimeric promoter MPr1112. The
CC promoter is derived from the petE promoter from pea plastocyanin gene,
CC by fusing the petE as-1 like and nos enhancer like elements to the
CC promoter MPr1098 (comprising TATA and CAAT boxes of petE), and then
CC fusing a fragment comprising a duplication of the element as2 and as1.
CC The petE promoter directs cell-specific but not full light-regulated
CC expression in transgenic tobacco plants. The promoter is used to
CC construct chimeric promoters of the invention. The specification
CC describes a chimeric expression promoter comprising a promoter of
CC the pea plastocyanin gene, or comprising a G box operably or
CC functionally linked upstream of a CAAT box, TATA box and transcription
CC initiation site. The chimeric promoters are used in expression vectors
CC for producing transgenic plants, such as dicotyledonous species,
CC e.g. potato, tobacco, cotton, lettuce, tomato, melon, cucumber, pea,
CC rape, beetroot, or sunflower, and monocotyledonous species, e.g. wheat,
CC barley, oat, rice, or corn.

XX Sequence 259 BP; 82 A; 67 C; 38 G; 72 T; 0 other;
SQ Query Match 94.9%; Score 75; DB 21; Length 259;
Best Local Similarity 100.0%; Pred.No. 2.4e-18;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 64
DB 13 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 72
QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87
RESULT 4
AAA96465
ID AAA96465 standard; DNA; 296 BP.
XX
AC AAA96465;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the promoter MP1111.
XX
KM petB promoter; chimeric promoter; transgenic plant; MP1111;
KW plastocyanin gene promoter; 88.
XX
OS Synthetic.
OS Pisum sativum.
XX
PN WO200056906-A1.
XX
PD 28-SEP-2000.
XX
PF 20-MAR-2000; 2000WO-IB00317.
XX
PR 22-MAR-1999; 99FR-0003635.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-587667/55.
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
PS Claim 2; Page 71; 83pp; English.
XX
CC The present sequence represents the chimeric promoter MP1111. The
CC promoter is derived from the petB promoter from pea plastocyanin gene,
CC by fusing the petB as-1 like and nos enhancer like elements to the
CC promoter MP1098 (comprising TATA and CAAT boxes of petB), and then
CC inserting a G box and fusing a fragment comprising a duplication of the
CC element as2 and as1. The petB promoter directs cell-specific but not
CC full light-regulated expression in transgenic tobacco plants. The
CC promoter is used to construct chimeric promoters of the invention. The
CC specification describes a chimeric expression promoter comprising a
CC promoter of the pea plastocyanin gene, or comprising a G box operably
CC or functionally linked upstream of a CAAT box, TATA box and transcription
CC initiation site. The chimeric promoters are used in expression vectors
CC for producing transgenic plants, such as dicotyledonous species,
CC e.g. potato, tobacco, cotton, lettuce, tomato, melon, cucumber, pea,
CC rape, beetroot, or sunflower, and monocotyledonous species, e.g. wheat,
CC barley, oat, rice, or corn.
XX
SQ Sequence 296 BP; 94 A; 74 C; 45 G; 83 T; 0 other;
Query Match 94.9%; Score 75; DB 21; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.5e-18;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 64
DB 13 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 72
QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87
RESULT 5
AAA96859
ID AAA96859 standard; DNA; 472 BP.
XX
AC AAA96859;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1169.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; 88.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 88; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 472 BP; 149 A; 92 C; 112 G; 119 T; 0 other;
Query Match 88.6%; Score 70; DB 21; Length 472;
Best Local Similarity 93.6%; Pred. No. 2.2e-16;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAA 60
DB 306 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAA 365
QY 61 AGGATGACGATGCCACT 78

Db 366 AGGATGACGATGCCAC 383

RESULT 6
AAA96858

ID AAA96858 standard; DNA; 541 BP.

XX AAA96858;

DT 19-FEB-2001 (first entry)

XX Nucleotide sequence of chimeric expression promoter Mp1168.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX W0200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000MO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

DR Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PT Claim 5, Page 87-88; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

XX Sequence 541 BP; 169 A; 104 C; 130 G; 138 T; 0 other;

XX Query Match 88.6%; Score 70; DB 21; Length 541;

XX Best Local Similarity 93.6%; Pred. NO. 2.2e-16;

XX Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATGCTGACACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60

Db 306 CTGCTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 365

QY 61 AGGATGACGATGCCAC 78

Db 366 AGGATGACGATGCCAC 383

RESULT 7

AAA96853

ID AAA96853 standard; DNA; 393 BP.

XX AAA96853;

AC 19-FEB-2001 (first entry)

XX Nucleotide sequence of chimeric expression promoter Mp1162.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX W0200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000MO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

DR Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PT Claim 5, Page 85; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

XX Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;

XX Query Match 88.1%; Score 69.6; DB 21; Length 393;

XX Best Local Similarity 94.7%; Pred. NO. 2.9e-16;

XX Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGCTGACACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 62

Db 227 TGGTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 286

QY 63 GGATGACGATGCCAC 78

Db 287 GGATGACGATGCCAC 302

RESULT 8

AAA96854

ID AAA96854 standard; DNA; 462 BP.

XX AAA96854;

XX 19-FEB-2001 (first entry)

XX Nucleotide sequence of chimeric expression promoter Mp1163.

KW Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99PR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PS WPI; 2000-647238/62.
XX
DR
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;
XX
Query Match 88.1%; Score 69.6; DB 21; Length 462;
Best Local Similarity 94.7%; Pred. No. 3e-16;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGTGACAGCTAGTGTGATGATGATCAAGATTGATGATTCCTCACTGACGTAAAG 62
DB 227 TGGTTACGACAGTGTGATGATGATCAAGATTGATGATTCCTCACTGACGTAAAG 286
QY 63 GGATGACGATGCCAC 78
DB 287 GGATGACGATGCCAC 302
XX
RESULT 9
AAA96856 standard; DNA; 600 BP.
ID AAA96856;
AC
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1165.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.

XX
XX WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99PR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PS WPI; 2000-647238/62.
XX
DR
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86-87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 600 BP; 188 A; 111 C; 147 G; 154 T; 0 other;
XX
Query Match 88.1%; Score 69.6; DB 21; Length 600;
Best Local Similarity 94.7%; Pred. No. 3.3e-16;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGTGACAGCTAGTGTGATGATGATCAAGATTGATGATTCCTCACTGACGTAAAG 62
DB 227 TGGTTACGACAGTGTGATGATGATCAAGATTGATGATTCCTCACTGACGTAAAG 286
QY 63 GGATGACGATGCCAC 78
DB 287 GGATGACGATGCCAC 302
XX
RESULT 10
AAA96855/c
ID AAA96855 standard; DNA; 392 BP.
AC
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1164.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX

PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 87.3%; Score 69; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 4.8e-16;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 10 GACTAGTATGATGTGATCAAGATTGATGATCTCCACTGACGTAAAGGATGAC 69
DB 302 GACTAGTATGATGTGATCAAGATTGATGATCTCCACTGACGTAAAGGATGAC 243
XX
QY 70 GCATGCCAC 78
DB 242 GCATGCCAC 234
XX
RESULT 11
AAA96857
ID AAA96857 standard; DNA; 604 BP.
XX
AC AAA96857;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1167.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
FN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.

XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 87; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
XX
Query Match 86.3%; Score 68.2; DB 21; Length 604;
Best Local Similarity 95.9%; Pred. No. 1.1e-15;
Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 6 TGACAGTATGATGTGATCAAGATTGATGATCTCCACTGACGTAAAGGA 65
DB 376 TGACAGTATGATGTGATCAAGATTGATGATCTCCACTGACGTAAAGGA 435
XX
QY 66 TGACGATGCCAC 78
DB 436 TGACGATGCCAC 448
XX
RESULT 12
AAA96839
ID AAA96839 standard; DNA; 371 BP.
XX
AC AAA96839;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1146.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
FN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 81; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comeline yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX
SQ Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;

Query Match 86.1%; Score 68; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 CTAGTATTGATGATATCAAGATTGATGATATCTCCAGCTAAGGATGACGC 71
DB 71 CTAGTATTGATGATATCAAGATTGATGATATCTCCAGCTAAGGATGACGC 130

QY 72 ATGCCACT 79
DB 131 ATGCCACT 138

RESULT 13
AAD03420/c
ID AAD03420 standard; DNA; 80 BP.

XX
XX AAD03420;
XX
DT 13-JUN-2001 (first entry)

DE CamV as-2/as-1 PCR primer #1 used for constructing MPr1130 promoter.
XX
XX Wheat; glutenin; transgenic plant; PRHMG-Dx5; MPr1130 promoter;
KM CamV 35S promoter; as-2 motif; as-1 motif; PCR primer; 88.

OS Cauliflower mosaic virus.
OS Synthetic.
XX
XX WO200123593-A1.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000MO-IB01383.
XX
XX 30-SEP-1999; 99FR-0012373.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
XX Gruber V, Norre F, Theisen M;
XX
XX WPI; 2001-245005/25.
XX
PT Chimeric expression promoters comprising nucleic acids from genes
PT encoding a high molecular weight wheat glutenin proteins, useful in the
PT production of transgenic plants -
XX
XX Example 3; Page 32; 120pp; English.

XX The patent discloses chimeric expression promoters comprising nucleic
XX acid sequences derived from the promoters of genes encoding high
XX molecular weight wheat glutenin proteins. The chimeric promoters
XX are used in standard recombinant DNA techniques for the generation
XX of transgenic plants and the production of polypeptides (e.g. marker
XX and therapeutic polypeptides).
XX The present DNA sequence is a PCR primer comprising the cauliflower
XX mosaic virus (CamV) 35S promoter as-2 motifs (in duplicate) and as-1
XX motif, used in the construction of the wheat high molecular weight

CC glutenin PRHMG-Dx5 promoter-derived chimeric promoter MPr1130 and
CC plasmids containing chimeric promoter sequences.
XX
SQ Sequence 80 BP; 21 A; 18 C; 16 G; 25 T; 0 other;

Query Match 73.4%; Score 58; DB 22; Length 80;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTATTGATGATATCAAGATTGATGATATCTCCAGCTAAGGATGACGCA 72
DB 58 GTATTGATGATATCAAGATTGATGATATCTCCAGCTAAGGATGACGCA 1

RESULT 14
AAD03396
ID AAD03396 standard; DNA; 236 BP.

XX
XX AAD03396;
XX
DT 13-JUN-2001 (first entry)

DE MPr1133 promoter DNA derived from wheat glutenin MPrHMG-Dx5 promoter.
XX
XX Wheat; glutenin; transgenic plant; MPr1133 promoter;
KM CamV 35S promoter; ds.
XX
XX Chimeric - Triticum aestivum.
OS Chimeric - Cauliflower mosaic virus.
XX
XX Key Location/Qualifiers
FH enhancer 78..132
FT /tag= a
FT /note= "As2/As2/As1 box from CamV 35S promoter"
FT misc_signal 198
FT /tag= b
FT /note= "transcription initiation site"

XX
XX WO200123593-A1.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000MO-IB01383.
XX
XX 30-SEP-1999; 99FR-0012373.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
XX Gruber V, Norre F, Theisen M;
XX
XX WPI; 2001-245005/25.
XX
PT Chimeric expression promoters comprising nucleic acids from genes
PT encoding a high molecular weight wheat glutenin proteins, useful in the
PT production of transgenic plants -
XX
XX Claim 4; Page 116; 120pp; English.

XX The patent discloses chimeric expression promoters comprising nucleic
XX acid sequences derived from the promoters of genes encoding high
XX molecular weight wheat glutenin proteins. The chimeric promoters
XX are used in standard recombinant DNA techniques for the generation
XX of transgenic plants and the production of polypeptides (e.g. marker
XX and therapeutic polypeptides).
XX The present DNA sequence is MPr1133 promoter which is derived by
XX deleting the sequence (comprising 2 prolamine-like boxes, 2 GATA boxes,
XX G-box and activating element) located upstream of the nucleotide -197
XX of MPr1130 promoter (AAD03394). The MPr1130 promoter is derived from
XX wheat high molecular weight glutenin PRHMG-Dx5 promoter DNA (AAD03390)
XX and contains elements from the cauliflower mosaic virus (CamV) 35S
XX promoter.
XX Note: The present sequence is referred as SEQ ID NO: 7 throughout the
XX specification. However this sequence is shown as SEQ ID NO: 35 in

CC the sequence listing.
XX
SQ Sequence 236 BP; 64 A; 70 C; 40 G; 62 T; 0 other;
Query Match 73.4%; Score 58; DB 22; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 15 GTGATTGATGATATCAAGATTGATATCTCCACGTAAGGATGACCGA 72
DB 75 GTGATTGATGATATCAAGATTGATATCTCCACGTAAGGATGACCGA 132
RESULT 15
AAD03397
ID AAD03397 standard; DNA; 299 BP.
AC AAD03397;
XX
DT 13-JUN-2001 (first entry)
XX
DE Mpr1134 promoter DNA derived from wheat glutenin MprHMG-Dx5 promoter.
XX
KW Wheat; glutenin; transgenic plant; Mpr1134 promoter;
KM CAMV 35S promoter; ds.
XX
OS Chimeric - Triticum aestivum.
OS Chimeric - Cauliflower mosaic virus.
XX
XX
Key Location/Qualifiers
FH 20..57
FT enhancer /*tag= a
FT enhancer 141..195
FT /*tag= b
FT /*note= "Aa2/Aa2/As1 box from CAMV 35S promoter"
FT misc_signal 261
FT /*tag= C
FT /*note= "Transcription initiation site"
XX
PN WO200123593-A1.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-IB01383.
XX
PR 30-SEP-1999; 99FR-0012373.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Gruber V, Norre F, Theisen M;
XX WPI; 2001-245005/25.
DR
XX
PT Chimeric expression promoters comprising nucleic acids from genes
PT encoding a high molecular weight wheat glutenin proteins, useful in the
PT production of transgenic plants -
XX
XX
PS Claim 4; Page 116; 120pp; English.
XX
XX The patent discloses chimeric expression promoters comprising nucleic
XX acid sequences derived from the promoters of genes encoding high
XX molecular weight wheat glutenin proteins. The chimeric promoters
XX are used in standard recombinant DNA techniques for the generation
XX of transgenic plants and the production of polypeptides (e.g. marker
XX and therapeutic polypeptides).
XX The present DNA sequence is Mpr1134 promoter which is derived by
XX deleting the sequence (comprising 2 prolamine-like boxes, 2 GATA boxes
XX and G-1-like box) located upstream of the nucleotide -260 of Mpr1130
XX promoter (AAD03394). The Mpr1130 promoter is derived from wheat high
XX molecular weight glutenin PrrHMG-Dx5 promoter DNA (AAD03390)
XX and contains elements from the cauliflower mosaic virus (CaMV) 35S
XX promoter.
XX Note: The present sequence is referred as SEQ ID NO: 8 throughout the

CC specification. However this sequence is shown as SEQ ID NO: 36 in
CC the sequence listing.
XX
SQ Sequence 299 BP; 78 A; 87 C; 49 G; 85 T; 0 other;
Query Match 73.4%; Score 58; DB 22; Length 299;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 15 GTGATTGATGATATCAAGATTGATATCTCCACGTAAGGATGACCGA 72
DB 138 GTGATTGATGATATCAAGATTGATATCTCCACGTAAGGATGACCGA 195
Search completed: May 11, 2003, 03:04:04
Job time : 154.644 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 836.727 seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79

Sequence: 1 catgcgcgcagctatgctat.....aaggatgcagctatgcctact 79

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: GenEmbl:

1: gb_ba:

2: gb_htg:

3: gb_in:

4: gb_ov:

5: gb_ph:

6: gb_pl:

7: gb_pr:

8: gb_ro:

9: gb_sts:

10: gb_sy:

11: gb_un:

12: gb_vl:

13: gb_vr:

14: em_ba:

15: em_fun:

16: em_hum:

17: em_in:

18: em_mu:

19: em_or:

20: em_ov:

21: em_ph:

22: em_pl:

23: em_ro:

24: em_sts:

25: em_sy:

26: em_un:

27: em_vl:

28: em_vr:

29: em_ba:

30: em_fun:

31: em_hum:

32: em_in:

33: em_mu:

34: em_or:

35: em_ov:

36: em_ph:

37: em_pl:

38: em_ro:

39: em_sts:

40: em_sy:

41: em_un:

42: em_vl:

43: em_vr:

44: em_ba:

45: em_fun:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	79	6 AX036611	AX036611 Sequence
2	79	100.0	79	6 AX036748	AX036748 Sequence
3	75	94.9	259	6 AX036616	AX036616 Sequence
4	75	94.9	296	6 AX036603	AX036603 Sequence
5	70	88.6	472	6 AX036759	AX036759 Sequence
6	70	88.6	541	6 AX036758	AX036758 Sequence
7	69.6	88.1	393	6 AX036753	AX036753 Sequence
8	69.6	88.1	462	6 AX036754	AX036754 Sequence
9	69.6	88.1	462	6 AX036756	AX036756 Sequence
10	69.6	88.1	9285	6 AX036757	AX036757 Sequence
11	69.6	88.1	15077	6 AX036755	AX036755 Sequence
12	69.6	88.1	392	6 AX036753	AX036753 Sequence
13	68.2	86.3	604	6 AX036757	AX036757 Sequence
14	68.2	86.3	371	6 AX036739	AX036739 Sequence
15	58	73.4	80	6 AX103782	AX103782 Sequence
16	58	73.4	236	6 AX103783	AX103783 Sequence
17	58	73.4	299	6 AX103784	AX103784 Sequence
18	58	73.4	332	6 AX103755	AX103755 Sequence
19	58	73.4	472	6 AX103753	AX103753 Sequence
20	55	69.6	80	6 AX103776	AX103776 Sequence
21	46	58.2	63	6 AX036610	AX036610 Sequence
22	46	58.2	63	6 AX036747	AX036747 Sequence
23	46	58.2	280	6 AX036601	AX036601 Sequence
24	46	58.2	301	6 AX036741	AX036741 Sequence
25	46	58.2	398	6 AX036740	AX036740 Sequence
26	39.2	49.6	63	6 AX103781	AX103781 Sequence
27	39.2	49.6	219	6 AX103756	AX103756 Sequence
28	39.2	49.6	282	6 AX103757	AX103757 Sequence
29	39.2	49.6	315	6 AX103758	AX103758 Sequence
30	39.2	49.6	381	6 AX103753	AX103753 Sequence
31	39.2	49.6	455	6 AX103754	AX103754 Sequence
32	39.2	49.6	505	6 AX103759	AX103759 Sequence
33	39	49.4	174	8 AF434749	AF434749 Zea mays
34	39	49.4	189	8 AF164073	AF164073 Sequence
35	39	49.4	197	8 AF434747	AF434747 Zea mays
36	39	49.4	199	8 AF434746	AF434746 Zea mays
37	39	49.4	199	8 AF434748	AF434748 Zea mays
38	39	49.4	199	8 AF434750	AF434750 Zea mays
39	39	49.4	240	6 AX033493	AX033493 Sequence
40	39	49.4	309	6 AX044092	AX044092 Sequence
41	39	49.4	309	6 AX405116	AX405116 Sequence
42	39	49.4	331	6 BD001990	BD001990 A transge
43	39	49.4	332	6 E01311	E01311 Cauliflower
44	39	49.4	333	6 I04847	I04847 Sequence 3
45	39	49.4	354	12 ARCAWPR	X04879 Canv Promot

ALIGNMENTS

RESULT 1	AX036611	79 bp	DNA	1linear	PAT 16-NOV-2000
LOCUS	AX036611				
DEFINITION	Sequence 17 from Patent FR2791358.				
ACCESSION	AX036611				
VERSION	AX036611.1	GI:11226206			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Rance, I., Theisen, M. and Gruber, V.				
JOURNAL	Patent: FR 2791358-A 17 29-SEP-2000;				
FEATURES	MERISTEM THERAPEUTICS (FR)				
	Location/Qualifiers				

Pred. No. is the number of results predicted by chance to have a

source 1..79
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional building block S7-Directional building block oligonucleotide for the construction of promoters by 1b-PCR"

BASE COUNT 22 a 15 c 20 g 22 t

Query Match 100.0%; Score 79; DB 6; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTGACAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60
DB 1 CATCTGACAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60

QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79

RESULT 2
AX036748
LOCUS AX036748 79 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 14 from Patent WO0058485.
ACCESSION AX036748
VERSION AX036748.1 GI:11226257
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 79)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 14 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1..79
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S7"

BASE COUNT 22 a 15 c 20 g 22 t

Query Match 100.0%; Score 79; DB 6; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTGACAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60
DB 1 CATCTGACAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60

QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79

RESULT 3
AX036616 259 bp DNA linear PAT 16-NOV-2000
LOCUS AX036616
DEFINITION Sequence 22 from Patent FR2791358.
ACCESSION AX036616
VERSION AX036616.1 GI:11226211
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 259)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
JOURNAL Patent: FR 2791358-A 22 29-SEP-2000;

FEATURES MERISTEM THERAPEUTICS (FR)
source location/Qualifiers
1..259
/organism="synthetic construct"
/db_xref="taxon:32630"
promoter 1..259
/note="Promoter MPr1112 differs from MPr111 by a deletion of 35 bp containing the 'G' box and stretching from position -127 to position -89 and a deletion of two bp situated at positions -78 and -76
promoter MPr1112"

BASE COUNT 82 a 67 c 38 g 72 t

Query Match 94.9%; Score 75; DB 6; Length 259;
Best Local Similarity 100.0%; Pred. No. 3.4e-16;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTAAGG 64
DB 13 CTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTAAGG 72

QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87

RESULT 4
AX036603 296 bp DNA linear PAT 16-NOV-2000
LOCUS AX036603
DEFINITION Sequence 9 from Patent FR2791358.
ACCESSION AX036603
VERSION AX036603.1 GI:11226198
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 296)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
JOURNAL Patent: FR 2791358-A 9 29-SEP-2000;
MERISTEM THERAPEUTICS (FR)
FEATURES
source location/Qualifiers
1..296
/organism="synthetic construct"
/db_xref="taxon:32630"
promoter 1..296
/note="Promoter MPr1111 created by inserting at -99 bp position of MPr1098, an 18 bp element containing a 'G' box and fusing a sequence of 58 bp (duplication of the element a62 and a61)
promoter MPr1111"

BASE COUNT 94 a 74 c 45 g 83 t

Query Match 94.9%; Score 75; DB 6; Length 296;
Best Local Similarity 100.0%; Pred. No. 3.4e-16;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTAAGG 64
DB 13 CTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCAGTACGTAAGG 72

QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87

RESULT 5
AX036759 472 bp DNA linear PAT 16-NOV-2000
LOCUS AX036759
DEFINITION Sequence 25 from Patent WO0058485.
ACCESSION AX036759
VERSION AX036759.1 GI:11226268

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 472)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .472
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1169"

149 a 92 c 112 g 119 t

Query Match
Best Local Similarity 88.6%; Score 70; DB 6; Length 472;
Pred. No. 2.1e-14;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATCTGACAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
DB 306 CTGTGTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 365
QY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 6
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

AX036758
Sequence 24 from Patent WO0058485.
AX036758
AX036758.1 GI:11226267
541 bp DNA linear PAT 16-NOV-2000

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 541)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 24 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .541
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1168"

169 a 104 c 130 g 138 t

Query Match
Best Local Similarity 88.6%; Score 70; DB 6; Length 541;
Pred. No. 2.1e-14;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATCTGACAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
DB 306 CTGTGTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 365
QY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 7

AX036753
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

AX036753
Sequence 19 from Patent WO0058485.
AX036753
AX036753.1 GI:11226262
393 bp DNA linear PAT 16-NOV-2000

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 393)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 19 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1162"

128 a 75 c 93 g 97 t

Query Match
Best Local Similarity 88.1%; Score 69.6; DB 6; Length 393;
Pred. No. 3e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 62
DB 227 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 286
QY 63 GGATGACGATGCCAC 78
DB 287 GGATGACGATGCCAC 302

RESULT 8
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

AX036754
Sequence 20 from Patent WO0058485.
AX036754
AX036754.1 GI:11226263
462 bp DNA linear PAT 16-NOV-2000

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 462)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 20 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .462
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1163"

148 a 87 c 111 g 116 t

Query Match
Best Local Similarity 88.1%; Score 69.6; DB 6; Length 462;
Pred. No. 2.9e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 62
DB 227 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 286
QY 63 GGATGACGATGCCAC 78

Db 287 GGATGACGCATGCCAC 302

RESULT 9
LOCUS AX036756 600 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 22 from Patent WO0058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS

SOURCE
ORGANISM
synthetic construct.
artificial sequences.
1 (bases 1 to 600)
REFERENCE
AUTHORS Rance, I., Theisen, M., and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 22 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Location/Qualifiers
1..600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MPRI165"
BASE COUNT 188 a 111 c 147 g 154 t
ORIGIN

Query Match 88.1%; Score 69.6; DB 6; Length 600;
Best Local Similarity 94.7%; Pred. No. 2.9e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGCAGCAGCTAGTGTGATGATATCAAGATGATGATCTCCACTGACGTAAG 62
Db 227 TGGTACGACTAGGATGATGATGATCAAGATGATGATCTCCACTGACGTAAG 286

Qy 63 GGATGACGCATGCCAC 78
Db 287 GGATGACGCATGCCAC 302

RESULT 10
LOCUS AX093047 9285 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 52 from Patent WO0118192.
ACCESSION AX093047
VERSION AX093047.1 GI:13509522
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
1 (bases 1 to 9285)
REFERENCE
AUTHORS Gruber, V., and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods for obtaining them
JOURNAL Patent: WO 0118192-A 52 15-MAR-2001;
MERISTEM THERAPEUTICS (FR)
FEATURES
Location/Qualifiers
1..9285
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="pMRT1336"
misc_feature
1
/note="pMRT1336 results from the insertion into pMRT1196 of the promoter MPRI165 isolated from plasmid pMRT1322 as described in PCT patent application PCT/IB00/00370"

BASE COUNT 2440 a 2252 c 2506 g 2087 t
ORIGIN

Query Match 88.1%; Score 69.6; DB 6; Length 9285;
Best Local Similarity 94.7%; Pred. No. 2.8e-14;

Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGCAGCAGCTAGTGTGATGATATCAAGATGATGATCTCCACTGACGTAAG 62
Db 6016 TGGTACGACTAGGATGATGATGATCAAGATGATGATCTCCACTGACGTAAG 6075

Qy 63 GGATGACGCATGCCAC 78
Db 6076 GGATGACGCATGCCAC 6091

RESULT 11
LOCUS AX093052 15077 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 57 from Patent WO0118192.
ACCESSION AX093052
VERSION AX093052.1 GI:13509527
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
1 (bases 1 to 15077)
REFERENCE
AUTHORS Gruber, V., and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods for obtaining them
JOURNAL Patent: WO 0118192-A 57 15-MAR-2001;
MERISTEM THERAPEUTICS (FR)
FEATURES
Location/Qualifiers
1..15077
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="pMRT1342"
misc_feature
1
/note="pMRT1342 results from the replacement of the expression cassette eP35S-gus-polyA35S from pMRT1335 by the expression cassette U5-gus-polyA35S isolated from pMRT1336"

BASE COUNT 3672 a 3892 c 4225 g 3288 t
ORIGIN

Query Match 88.1%; Score 69.6; DB 6; Length 15077;
Best Local Similarity 94.7%; Pred. No. 2.7e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGCAGCAGCTAGTGTGATGATATCAAGATGATGATCTCCACTGACGTAAG 62
Db 7055 TGGTACGACTAGGATGATGATGATCAAGATGATGATCTCCACTGACGTAAG 7114

Qy 63 GGATGACGCATGCCAC 78
Db 7115 GGATGACGCATGCCAC 7130

RESULT 12
LOCUS AX036755 392 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
1 (bases 1 to 392)
REFERENCE
AUTHORS Rance, I., Theisen, M., and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Location/Qualifiers
1..392
/organism="synthetic construct"

	Promoter	127 a	80 c	87 g	98 t	/db_xref="taxon:32630" /note="promoter MPRI164"
BASE COUNT ORIGIN						1..392
Query Match	Best Local Similarity	87.3%;	Score 69;	DB 6;	Length 392; Matches 69; Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
OY	10	GACTGATGTATTGATGGTATATCAAGTTTGATCTCCACCTGACGTAAAGGATGC	69			
Db	302	GACTGATGTATTGATGGTATTCAGAATTGATGTAATCTCCACCTGAAGGATGAC	243			
OY	70	GCATGCCAC	78			
Db	242	GCATGCCAC	234			
RESULT 13						
LOCUS	AXO36757	604 bp	DNA	linear	PAT 16-NOV-2000	
DEFINITION	A sequence 23 from Patent WO058485.					
ACCESSION	AXO36757					
VERSION	AXO36757.1 GI:11226266					
KEYWORDS	Synthetic construct. Synthetic construct artificial sequences.					
SOURCE	Rance,I., Theisen,M. and Gruber,V. Chimeric expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus Patent: WO 0058485-A 23 05-OCT-2000; MERISTEM TERAPEUTICS (FR) ; RANGE INAN (FR) ; THEISEN MANFRED (FR) GRUBER VERONIQUE (FR)					
REFERENCE						
AUTHORS	Title					
JOURNAL						
FEATURES	Source	Location/Qualifiers				
	1..604	organism="synthetic construct"				
	/db_xref="taxon:32630"					
	/note="Promoter MPR1167"					
BASE COUNT	Promoter	186 a	116 c	145 g	157 t	
ORIGIN						
Query Match	Best local similarity	86.3%;	Score 68.2;	DB 6;	Length 604; Matches 70; Conservative 0;	Mismatches 3; Indels 0; Gaps 0;
OY	6	TSCAGACTGATGATGATGATGATCATCAGATTTGATGTAATCTCCACTGACGTAAAGGA	65			
Db	376	TSCAGACTGATGATGATGATGATGATCATCAGATTTGATGTAATCTCCACTGACGTAAAGGA	435			
OY	66	TGACCGATGCCAC	78			
Db	436	TGACCGATGCCAC	448			
RESULT 14						
LOCUS	AXO36739	371 bp	DNA	linear	PAT 16-NOV-2000	
DEFINITION	A sequence 5 from Patent WO058485.					
ACCESSION	AXO36739					
VERSION	AXO36739.1 GI:11226248					
KEYWORDS	Synthetic construct. Synthetic construct artificial sequences.					
SOURCE	Rance,I., Theisen,M. and Gruber,V. Chimeric expression promoters originating from commelina yellow mottle virus and Cassava vein mosaic virus Patent: WO 0058485-A 5 05-OCT-2000;					
REFERENCE						
AUTHORS	Title					
JOURNAL						

```

MEISTERHEIM THERAPEUTICS (FR) ; RANGE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
    source          Location/Qualifiers
                    1..371
                    /organism="synthetic construct"
                    /db_xref="taxon:32630"
                    /note="promoter MPr1146"
BASE COUNT        122 a      68 c      89 g      92 t
ORIGIN
    Query Match      86.1%; Score 68; DB 6; Length 371;
    Best Local Similarity 100.0%; Pred. No. 1.1e-13;
    Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy 12 CTAGATTGATGTGATATCAAGATTGATGTGATATCTCCACTGACGTAAAGGATGACGC 71
    Db 71 CTAGATTGATGTGATATCAAGATTGATGTGATATCTCCACTGACGTAAAGGATGACGC 130
    Oy 72 ATGCCACT 79
    Db 131 ATGCCACT 138

RESULT 15
LOCUS              AXI03782/c      80 bp      DNA      linear      PAT 30-APR-2001
DEFINITION         Sequence 34 from Patent WO0123593.
ACCESSION          AXI03782
VERSION            AXI03782.1  GI:13919991
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS            Gruber,V., Norre,F. and Theisen,M.
TITLE              Synthetic and chimeric promoters, expression cassettes, plasmids,
                  vectors, transgenic plants and seeds containing them, and method
                  for producing them
                  Patent: WO 0123593-A 34 05-APR-2001;
                  Meristem Therapeutics Intellectual Property (FR)
JOURNAL
FEATURES
    source          Location/Qualifiers
                    1..80
                    /organism="synthetic construct"
                    /db_xref="taxon:32630"
                    /note="Description of Artificial
                    Sequence:Oligodeoxynucleotide"
BASE COUNT        21 a      18 c      16 g      25 t
ORIGIN
    Query Match      73.4%; Score 58; DB 6; Length 80;
    Best Local Similarity 100.0%; Pred. No. 4.5e-10;
    Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy 15 GTGATTGATGTGATATCAAGATTGATGTGATATCTCCACTGACGTAAAGGATGACGC 72
    Db 58 GTGATTGATGTGATATCAAGATTGATGTGATATCTCCACTGACGTAAAGGATGACGC 1

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 31.6792 Seconds

(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgcgcagactgactgatt.....aaggatgacatgcact 79

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	49.4	439	1	US-08-247-809A-3
2	39	49.4	439	2	US-08-711-728-3
3	39	49.4	446	1	US-08-764-100-23
4	39	49.4	532	3	US-09-042-426-1
5	39	49.4	532	4	US-09-281-238-1
6	39	49.4	532	4	US-09-330-760-1
7	39	49.4	532	4	US-09-328-473-1
8	39	49.4	532	4	US-09-330-737-1
9	39	49.4	532	4	US-09-329-169-1
10	39	49.4	532	4	US-09-330-714A-1
11	39	49.4	532	4	US-09-328-826-1
12	39	49.4	560	4	US-09-042-426-5
13	39	49.4	560	4	US-09-291-238-5
14	39	49.4	560	4	US-09-330-760-5
15	39	49.4	560	4	US-09-328-473-5
16	39	49.4	560	4	US-09-330-737-5
17	39	49.4	560	4	US-09-329-169-5
18	39	49.4	560	4	US-09-330-714A-5
19	39	49.4	560	4	US-09-328-826-5
20	39	49.4	661	4	US-09-027-998A-33
21	39	49.4	793	4	US-08-371-764-1
22	39	49.4	793	4	US-08-897-736-1
23	39	49.4	831	1	US-08-450-834-5
24	39	49.4	950	6	5177308-3
25	39	49.4	978	1	US-08-446-486-31
26	39	49.4	978	1	US-08-463-308-31
27	39	49.4	979	1	US-08-446-486-30

28	39	49.4	979	1	US-08-463-308-30	Sequence 30, Appl
29	39	49.4	980	6	5254799-30	Patent No. 5254799
30	39	49.4	1030	4	US-07-936-163-46	Sequence 46, Appl
31	39	49.4	1030	4	US-08-729-601A-43	Sequence 43, Appl
32	39	49.4	1034	4	US-09-363-970-35	Sequence 35, Appl
33	39	49.4	1138	4	US-09-011-151-8	Sequence 8, Appl
34	39	49.4	1138	4	US-09-011-151-9	Sequence 9, Appl
35	39	49.4	1136	4	US-08-729-601A-46	Sequence 46, Appl
36	39	49.4	1279	4	US-09-185-244-2	Sequence 2, Appl
37	39	49.4	1279	4	US-09-471-813-6	Sequence 6, Appl
38	39	49.4	1303	3	US-08-894-440-2	Sequence 2, Appl
39	39	49.4	1303	4	US-09-458-093-2	Sequence 2, Appl
40	39	49.4	1651	3	US-09-065-999-5	Sequence 5, Appl
41	39	49.4	1651	3	US-09-065-999-6	Sequence 6, Appl
42	39	49.4	1722	1	US-08-247-809A-5	Sequence 5, Appl
43	39	49.4	1722	2	US-08-711-728-5	Sequence 5, Appl
44	39	49.4	1829	1	US-07-966-187-17	Sequence 17, Appl
45	39	49.4	1863	1	US-08-525-507-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-247-809A-3
; Sequence 3, Application US/08247809A
; Patent No. 5569823
GENERAL INFORMATION:
APPLICANT: Peter H. Schreier, Klaus Stenzel, Gunter Adam;
APPLICANT: Edgar Mals
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSER: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
MEDIUM TYPE: storage
COMPUTER: NEC Powermate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,809A
FILING DATE: May 23, 1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 178 45.6 (Germany)
FILING DATE: May 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9049-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-247-809A-3

Query Match 49.4%; Score 39; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 6-2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 338 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 376

RESULT 2

US-08-711-728-3
Sequence 3, Application US/08711728
Patent No. 5973135
GENERAL INFORMATION:
APPLICANT: Peter H. Schreier, Klaus Stenzel, Gunter Adam;
APPLICANT: Edgar Maier
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: NEC Powermate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711.728
FILING DATE: 03-SEPT-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,809
FILING DATE: 23-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 43178456
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9049.1-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-711-728-3

Query Match 49.4%; Score 39; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 338 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 376

RESULT 3

US-08-764-100-23
Sequence 2, Application US/08764100
Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus O.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielens L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.

TITLE OF INVENTION: Improvements in or Relating to Organic
NUMBER OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 577370018, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 354-1125
INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 446 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-23

Query Match 49.4%; Score 39; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 4

US-09-042-426-1
Sequence 1, Application US/09042426
Patent No. 6114608
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6114608artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION/DOCKET NUMBER: 32,993
TELECOMMUNICATION INFORMATION: 135/1
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-042-426-1

Query Match 49.4%; Score 39; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 312

RESULT 5
US-09-291-238-1
Sequence 1, Application US/09291238
Patent No. 6222104
GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6222104artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/291.238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-291-238-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 312

RESULT 6
US-09-330-760-1
Sequence 1, Application US/09330760
Patent No. 6229075
GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6229075artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,760
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-760-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCGA 312

RESULT 7
US-09-328-473-1
Sequence 1, Application US/09328473
Patent No. 6232533

```

GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 623253artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328,473
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-328-473-1

Query Match          49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 8
US-09-330-737-1
Sequence 1, Application US/09330737
Patent No. 6232534
GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 623253artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/09/330,737
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-737-1

Query Match          49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 9
US-09-329-169-1
Sequence 1, Application US/09329169
Patent No. 6329575
GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 632957artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,169
FILING DATE: 09-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
```

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-329-169-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 10
US-09-330-714A-1
Sequence 1, Application US/09330714A
Patent No. 6342660
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
Thuringiensis Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6342660artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330.714A
FILING DATE: 11-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042.426
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-330-714A-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72

Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 11
US-09-328-826-1
Sequence 1, Application US/09328826
Patent No. 6399860
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6399860artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328.826
FILING DATE: 09-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/042.426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-328-826-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 12
US-09-042-426-5
Sequence 5, Application US/09042426
Patent No. 6114608
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6114608artis Corporation
STREET: 564 Morris Avenue
CITY: Summit

STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-042-426-5

Query Match 49.4%; Score 39; DB 3; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 13
US-09-291-238-5
Sequence 5, Application US/09291238
Patent No. 6222104
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6222104artls Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/291,238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-760-5

Query Match 49.4%; Score 39; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-291-238-5

Query Match 49.4%; Score 39; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 14
US-09-330-760-5
Sequence 5, Application US/09330760
Patent No. 6229075
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6229075artls Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,760
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-760-5

Query Match 49.4%; Score 39; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 322 GATTGATGTATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 15
 ; US-09-328-473-5
 ; Sequence 5, Application US/09328473
 ; Patent No. 6232533
 ; GENERAL INFORMATION:
 ; APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
 ; TITLE OF INVENTION: DNA Construct Containing Bacillus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6232533artis Corporation
 ; STREET: 564 Morris Avenue
 ; CITY: Summit
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07901
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/328,473
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/042,426
 ; FILING DATE: March 13, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hoxie, Thomas
 ; REGISTRATION NUMBER: 32,993
 ; REFERENCE/DOCKET NUMBER: 135/1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (919) 541-8614
 ; TELEFAX: (919) 541-8689
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 560 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; IMMEDIATE SOURCE:
 ; CLONE: 358 Promoter
 ; US-09-328-473-5

Query Match 49.4%; Score 39; DB 4; Length 560;
 Best Local Similarity 100.0%; Pred. No. 6,7e-06;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 322 GATTGATGTATATCTCCACTGACGTAAAGGATGACGCA 360

Search completed: May 11, 2003, 03:07:42;
 Job time : 33.6792 secs

US-09-963-803-5
; Sequence 5, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1146
; NAME/KEY: promoter
; LOCATION: (1)..(371)
; OTHER INFORMATION:
US-09-963-803-5

Query Match 86.1%; Score 68; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 3,7e-15;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 CTAGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGC 71
Db 71 CTAGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGC 130
|||
Qy 72 ATGCCACT 79
|||
Db 131 ATGCCACT 138

RESULT 10
US-09-870-375-34/c
; Sequence 34, Application US/09870375
; Patent No. US20020083486A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
; FILE REFERENCE: PTHMWG1
; CURRENT APPLICATION NUMBER: US/09/870,375
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 80
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence:Oligodeoxynucleotide
US-09-870-375-34

Query Match 73.4%; Score 58; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 9,3e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 GTGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGCA 72
Db 58 GTGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGCA 1

RESULT 11
US-09-870-375-35
; Sequence 35, Application US/09870375
; Patent No. US20020083486A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
; FILE REFERENCE: PTHMWG1
; CURRENT APPLICATION NUMBER: US/09/870,375
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MP1133
; OTHER INFORMATION: promoter
; NAME/KEY: misc_feature
; LOCATION: (78)..(132)
; OTHER INFORMATION: As2/As2/As1 box
; NAME/KEY: misc_feature
; LOCATION: (198)
; OTHER INFORMATION: Transcription Initiation Site
US-09-870-375-35

Query Match 73.4%; Score 58; DB 10; Length 236;
Best Local Similarity 100.0%; Pred. No. 1,3e-11;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 GTGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGCA 72
Db 75 GTGATTGATGATGATCAAGATTGATGATCTCCACGTAAGGATGACGCA 132
|||

RESULT 12
US-09-870-375-36
; Sequence 36, Application US/09870375
; Patent No. US20020083486A1
; GENERAL INFORMATION:
; APPLICANT: MERISTEM THERAPEUTICS
; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
; FILE REFERENCE: PTHMWG1
; CURRENT APPLICATION NUMBER: US/09/870,375
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MP1134
; NAME/KEY: enhancer
; LOCATION: (20)..(57)
; OTHER INFORMATION: Enhancer box
; NAME/KEY: misc_feature
; LOCATION: (141)..(195)
; OTHER INFORMATION: As2/As2/As1 box
; NAME/KEY: misc_feature
; LOCATION: (261)
; OTHER INFORMATION: Transcription Initiation Site
US-09-870-375-36

Query Match 73.4%; Score 58; DB 10; Length 299;
Best Local Similarity 100.0%; Pred. No. 1,5e-11;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 138 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 195

RESULT 13

US-09-870-375-7
 ; Sequence 7, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 332
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MP1135
 ; NAME/KEY: misc feature
 ; LOCATION: (21)..(28)
 ; OTHER INFORMATION: G-like box
 ; NAME/KEY: enhancer
 ; LOCATION: (53)..(90)
 ; OTHER INFORMATION: Enhancer box
 ; NAME/KEY: misc feature
 ; LOCATION: (174)..(228)
 ; OTHER INFORMATION: As2/As2/As1 box
 ; NAME/KEY: TATA signal
 ; LOCATION: (264)..(270)
 ; OTHER INFORMATION: TATA box
 ; NAME/KEY: misc feature
 ; LOCATION: (294)
 ; OTHER INFORMATION: Transcription Initiation Site
 ; US-09-870-375-7

Query Match 73.4%; Score 58; DB 10; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.5e-11;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 171 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 228

RESULT 14

US-09-870-375-5
 ; Sequence 5, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 472
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MP1130
 ; OTHER INFORMATION: promoter

NAME/KEY: misc feature
 ; LOCATION: (22)..(29)
 ; OTHER INFORMATION: Proline-1-like box
 ; NAME/KEY: misc feature
 ; LOCATION: (70)..(73)
 ; OTHER INFORMATION: GATA box
 ; NAME/KEY: misc feature
 ; LOCATION: (87)..(90)
 ; OTHER INFORMATION: GATA box
 ; NAME/KEY: misc feature
 ; LOCATION: (127)..(133)
 ; OTHER INFORMATION: Proline-1-like box
 ; NAME/KEY: misc feature
 ; LOCATION: (161)..(168)
 ; OTHER INFORMATION: G-like box
 ; NAME/KEY: enhancer
 ; LOCATION: (193)..(230)
 ; OTHER INFORMATION: Enhancer box
 ; NAME/KEY: misc feature
 ; LOCATION: (314)..(368)
 ; OTHER INFORMATION: As2/As2/As1 box
 ; NAME/KEY: TATA signal
 ; LOCATION: (404)..(410)
 ; OTHER INFORMATION: TATA box
 ; NAME/KEY: misc feature
 ; LOCATION: (434)
 ; OTHER INFORMATION: Transcription Initiation Site
 ; US-09-870-375-5

Query Match 73.4%; Score 58; DB 10; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.7e-11;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 311 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 368

RESULT 15

US-09-870-375-28
 ; Sequence 28, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 80
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial
 ; OTHER INFORMATION: Sequence:Oligodeoxynucleotide
 ; US-09-870-375-28

Query Match 69.6%; Score 55; DB 10; Length 80;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 ATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 1 ATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 55

Search completed: May 11, 2003, 06:11:43
 Job time : 79.208 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 360.11 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34
Sequence: 1 gactccctcactcagtcgctgagaca 34Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```
GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_str:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_str:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrc:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	34	AX036749	AX036749 Sequence
2	21.6	66.24	6	AC024376	AC024376 Homo sapi
3	21.4	62.9	10	AC091523	AC091523 Mus muscu
4	20.8	61.2	137875	AC123002	AC123002 Rattus no
5	20.8	61.2	178549	AC109962	AC109962 Rattus no
6	20.2	59.4	109939	AC009300	AC009300 Homo sapi
7	20.2	59.4	11081	AC010249	AC010249 Homo sapi
8	20.2	59.4	137165	AC011264	AC011264 Homo sapi
9	20.2	59.4	162918	AC026729	AC026729 Homo sapi
10	20.2	59.4	176172	AC121165	AC121165 Rattus no
11	20.2	59.4	185791	AC022438	AC022438 Homo sapi
12	20	58.8	80229	AC130200	AC130200 Medicago
13	20	58.8	159746	AC121899	AC121899 Mus muscu
14	20	58.8	170083	AC107106	AC107106 Rattus no
15	20	58.8	227789	AC129318	AC129318 Mus muscu
16	19.8	58.2	2539	BC026428	BC026428 Mus muscu
17	19.8	58.2	2887	BC024955	BC024955 Mus muscu
18	19.8	58.2	51309	AC104742	AC104742 Mus muscu
19	19.8	58.2	51309	AC104742	AC104742 Mus muscu
20	19.8	58.2	51343	AC108017	AC108017 Homo sapi
21	19.8	58.2	84270	AC114288	AC114288 Homo sapi
22	19.8	58.2	125164	AC115258	AC115258 Rattus no
23	19.8	58.2	136917	AC087220	AC087220 Oryza sat
24	19.8	58.2	144200	AC022638	AC022638 Homo sapi
25	19.8	58.2	147739	AC119106	AC119106 Rattus no
26	19.8	58.2	150119	AL845366	AL845366 Danio rer
27	19.8	58.2	152614	AC106797	AC106797 Homo sapi
28	19.8	58.2	166623	AC115858	AC115858 Mus muscu
29	19.8	58.2	168056	AC099254	AC099254 Rattus no
30	19.8	58.2	168946	AC114449	AC114449 Rattus no
31	19.8	58.2	185859	AC068026	AC068026 Homo sapi
32	19.6	57.6	10821	AE012340	AE012340 Xanthomon
33	19.6	57.6	60009	AC023436	AC023436 Homo sapi
34	19.6	57.6	69523	AC124305	AC124305 Homo sapi
35	19.6	57.6	118684	AC004537	AC004537 Homo sapi
36	19.6	57.6	118684	AC004259	AC004259 Human Chr
37	19.6	57.6	121963	AC018865	AC018865 Homo sapi
38	19.6	57.6	128931	AC107498	AC107498 Rattus no
39	19.6	57.6	133863	AC004600	AC004600 Homo sapi
40	19.6	57.6	159183	AC122031	AC122031 Mus muscu
41	19.6	57.6	159952	AC090976	AC090976 Bos tauru
42	19.6	57.6	171271	AC124311	AC124311 Homo sapi
43	19.6	57.6	173067	AC112749	AC112749 Rattus no
44	19.6	57.6	179113	AC025895	AC025895 Homo sapi
45	19.6	57.6	179933	AC124997	AC124997 Homo sapi

ALIGNMENTS

RESULT 1
AX036749
LOCUS AX036749 34 bp DNA
DEFINITION Sequence 15 from Patent WO0058485.
ACCESSION AX036749
VERSION AX036749.1 GI:11226258

KEYWORDS
ORGANISM
synthetic construct.
synthetic construct.
artificial sequences.

REFERENCE
1 (bases 1 to 34)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 15 05-OCT-2000;

FEATURES
SOURCE
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1..34
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Guide desoxynucleotide building block G1"

BASE COUNT
ORIGIN
8 a 9 c 7 g 10 t

Query Match 100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACTCCTACTTATGATCGTACTGAGACA 34
Db 1 GACTCCTACTTATGATCGTACTGAGACA 34

RESULT 2
AC024376 66624 bp DNA linear HTG 05-JAN-2001
LOCUS
DEFINITION Homo sapiens clone RP11-610N12, LOW-PASS SEQUENCE SAMPLING.
AC024376
AC024376.3 GI:12039551
VERSION
KEYWORDS HTG; HTGS PHASE0.
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 66624)
Britten,B., Lincon,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-610N12
Unpublished
2 (bases 1 to 66624)
Britten,B., Lincon,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Bogunlavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Casale,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galegan,J., Gaidyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largoccone,K., Lenoczky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., McDonald,P., Marguis,N., McCarthy,M.,
McBwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamae,J., Testaye,S., Theodore,J., Tirrell,A.,
Traversan,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 5, 2001 this sequence version replaced gi:1767813.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6257
Center clone name: 610_N_12

* NOTE: This record contains 82 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 711 810: contig of 710 bp in length
* 811 1530: contig of 720 bp in length
* 1531 1630: gap of 100 bp
* 1631 2328: contig of 698 bp in length
* 2328 2428: gap of 100 bp
* 2429 3164: contig of 736 bp in length
* 3165 3264: gap of 100 bp
* 3265 3971: contig of 707 bp in length
* 3972 4071: gap of 100 bp
* 4072 4770: contig of 699 bp in length
* 4771 4870: gap of 100 bp
* 4871 5593: contig of 723 bp in length
* 5594 5693: gap of 100 bp
* 5694 6416: contig of 723 bp in length
* 6417 6516: gap of 100 bp
* 6517 7230: contig of 714 bp in length
* 7231 7330: gap of 100 bp
* 7331 8041: contig of 711 bp in length
* 8042 8141: gap of 100 bp
* 8142 8862: contig of 721 bp in length
* 8863 8962: gap of 100 bp
* 8963 9693: contig of 731 bp in length
* 9694 9793: gap of 100 bp
* 9794 10506: contig of 713 bp in length
* 10507 10606: gap of 100 bp
* 10607 11318: contig of 712 bp in length
* 11319 11418: gap of 100 bp
* 11419 12132: contig of 714 bp in length
* 12133 12232: gap of 100 bp
* 12233 12950: contig of 718 bp in length
* 12951 13050: gap of 100 bp
* 13051 13772: contig of 722 bp in length
* 13773 13872: gap of 100 bp
* 13873 14584: contig of 712 bp in length
* 14585 14684: gap of 100 bp
* 14685 15399: contig of 715 bp in length
* 15400 15499: gap of 100 bp
* 15500 16197: contig of 698 bp in length
* 16198 16297: gap of 100 bp
* 16298 17002: contig of 705 bp in length
* 17003 17102: gap of 100 bp
* 17103 17825: contig of 723 bp in length
* 17826 17925: gap of 100 bp
* 17926 18645: contig of 720 bp in length
* 18646 18745: gap of 100 bp
* 18746 19459: contig of 714 bp in length
* 19460 19559: gap of 100 bp
* 19560 20279: contig of 720 bp in length
* 20280 20379: gap of 100 bp
* 20380 21081: contig of 702 bp in length
* 21082 21181: gap of 100 bp
* 21182 21887: contig of 706 bp in length
* 21888 21987: gap of 100 bp
* 21988 22699: contig of 712 bp in length
* 22700 22799: gap of 100 bp
* 22800 23519: contig of 720 bp in length
* 23520 23619: gap of 100 bp
* 23620 24366: contig of 747 bp in length
* 24367 24466: gap of 100 bp
* 24467 25185: contig of 719 bp in length
* 25186 25285: gap of 100 bp
* 25286 25989: contig of 704 bp in length

[illegible]

```

* 55295 56008: contig of 714 bp in length
* 56009 56108: gap of 100 bp

Query Match 63.5%; Score 21.6; DB 2; Length 66624;
Best Local Similarity 82.8%; Pred. No. 42;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GACTTCCTACTTATGCATCGCTACTG 29
|||||
Db 28524 GACTCCGCTACTGATGTATCGTACGGTG 28552
|||||

RESULT 3
AC091523
LOCUS AC091523 235505 bp DNA linear ROD 19-FEB-2002
DEFINITION Mus musculus chromosome 1 clone rp23-77a8, complete sequence.
ACCESSION AC091523
VERSION AC091523.22 GI:18702401
KEYWORDS HTG.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 235505)
Deschamps,S., Chambliss,D. and Roe,B.A.
Mus musculus Chromosome 1 BAC Clone rp23-77a8
Unpublished
2 (bases 1 to 235505)
Deschamps,S., Chambliss,D. and Roe,B.A.
Direct Submission
Submitted (29-APR-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 235505)
Deschamps,S., Chambliss,D. and Roe,B.A.
Direct Submission
Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
Source
1. 235505
/organism="Mus musculus"
/ld_xref="taxon:10090"
/chromosome="1"
/clone="rp23-77a8"
/clone_1lb="RPCI Mouse BAC Library 23"
BASE COUNT 69731 a 50058 c 49581 g 66135 t
ORIGIN
Query Match 62.9%; Score 21.4; DB 10; Length 235505;
Best Local Similarity 80.6%; Pred. No. 51;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CTCCTCTACTTATGCATCGCTACTGTGAGAC 33
|||||
Db 154904 CTCCTCTACTTATGCCTCAGTACGTGGCAC 154934
|||||

RESULT 4
AC123002/c
LOCUS AC123002 137875 bp DNA linear HTG 24-JUL-2002
DEFINITION Rattus norvegicus clone CH230-22H7, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC123002
VERSION AC123002.2 GI:21909223
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.

```

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 137875)
Muzny, D.M., Adams, C., Adio-Oduola, B., Al-Osman, F.R., Allen, C.,

Albrooke, S.L., Amaralunga, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonin, D.,
Bouch, J., Bowie, S., Bivleva, M., Brown, E., Brown, M., Bryant, N.P.,
Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dachorne, S.R., David, R.,
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, R.R., Delgado, O., Denn, A.L., Ding, Y., Dihn, H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Einhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Homes, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loussaged, H.,
Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Mabswat, M., Mapua, P., Martin, R., Matindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogburn, M., Okunolu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peterson, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Schierer, S., Scott, G., Shen, H., Shoochattari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sulton, A., Swalek, A., Tabor, P., Tameria, A., Tamerias, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vazquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL

Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 137875)
Worley, K.C.

JOURNAL

Submitted (26-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 137875)
Worley, K.C.

JOURNAL

Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 19, 2002 this sequence version replaced gi:21217336.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GYTF

Center clone name: CH230-22H7

Summary Statistics

Sequencing vector: pLasmid

Chemistry: Dye-terminator Big Dye: 1004 of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 89991 bases at least Q40

Consensus quality: 97350 bases at least Q30
Consensus quality: 101050 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently
consists of 63 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1084:	contig of 1084 bp in length
1085	1184:	gap of unknown length
1185	2985:	contig of 1801 bp in length
2986	3085:	gap of unknown length
3086	4594:	contig of 1509 bp in length
4595	4694:	gap of unknown length
4695	6070:	contig of 1376 bp in length
6071	6170:	gap of unknown length
6171	7704:	contig of 1534 bp in length
7705	7804:	gap of unknown length
7805	8984:	contig of 1180 bp in length
8985	9084:	gap of unknown length
9085	10222:	contig of 1138 bp in length
10223	10322:	gap of unknown length
10323	11526:	contig of 1204 bp in length
11527	11626:	gap of unknown length
11627	12655:	contig of 1029 bp in length
12656	12755:	gap of unknown length
12756	13837:	contig of 1082 bp in length
13838	13937:	gap of unknown length
13938	15178:	contig of 1241 bp in length
15179	15278:	gap of unknown length
15279	17550:	contig of 2272 bp in length
17551	17650:	gap of unknown length
17651	19171:	contig of 1521 bp in length
19172	19271:	gap of unknown length
19272	20695:	contig of 1424 bp in length
20696	20795:	gap of unknown length
20796	21925:	contig of 1130 bp in length
21926	22025:	gap of unknown length
22026	23502:	contig of 1477 bp in length
23503	23602:	gap of unknown length
23603	24946:	contig of 1344 bp in length
24947	25046:	gap of unknown length
25047	26306:	contig of 1260 bp in length
26307	26406:	gap of unknown length
26407	27933:	contig of 1527 bp in length
27934	28033:	gap of unknown length
28034	29051:	contig of 1018 bp in length
29052	29151:	gap of unknown length
29152	31029:	contig of 1878 bp in length
31030	31129:	gap of unknown length
31130	32138:	contig of 1009 bp in length
32139	32238:	gap of unknown length
32239	33644:	contig of 1406 bp in length
33645	33744:	gap of unknown length
33745	35209:	contig of 1465 bp in length
35210	35309:	gap of unknown length
35310	37128:	contig of 1819 bp in length
37129	37228:	gap of unknown length
37229	38472:	contig of 1244 bp in length
38473	38572:	gap of unknown length
38573	39697:	contig of 1125 bp in length
39698	39797:	gap of unknown length
39799	41826:	contig of 2029 bp in length
41827	41926:	gap of unknown length
41927	42965:	contig of 1039 bp in length
42966	43065:	gap of unknown length
43066	44598:	contig of 1533 bp in length
44599	44698:	gap of unknown length

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1158: contig of 1158 bp in length
 * 1159 1258: gap of unknown length
 * 1259 3027: contig of 1769 bp in length
 * 3028 3127: gap of unknown length
 * 3128 4598: contig of 1471 bp in length
 * 4599 4698: gap of unknown length
 * 4699 6401: contig of 1703 bp in length
 * 6402 6501: gap of unknown length
 * 6502 8270: contig of 1769 bp in length
 * 8271 8370: gap of unknown length
 * 8371 9753: contig of 1383 bp in length
 * 9754 9853: gap of unknown length
 * 9854 10988: contig of 1135 bp in length
 * 10989 11088: gap of unknown length
 * 11089 12826: contig of 1738 bp in length
 * 12827 12926: gap of unknown length
 * 12927 14308: contig of 1382 bp in length
 * 14309 14408: gap of unknown length
 * 14409 15711: contig of 1303 bp in length
 * 15712 15811: gap of unknown length
 * 15812 17817: contig of 2006 bp in length
 * 17818 17917: gap of unknown length
 * 17918 18963: contig of 1046 bp in length
 * 18964 19063: gap of unknown length
 * 19064 20236: contig of 1173 bp in length
 * 20237 20336: gap of unknown length
 * 20337 21839: contig of 1503 bp in length
 * 21840 21939: gap of unknown length
 * 21940 23344: contig of 1405 bp in length
 * 23345 23444: gap of unknown length
 * 23445 24472: contig of 1028 bp in length
 * 24473 24572: gap of unknown length
 * 24573 26191: contig of 1619 bp in length
 * 26192 26291: gap of unknown length
 * 26292 27405: contig of 1114 bp in length
 * 27406 27505: gap of unknown length
 * 27506 29085: contig of 1580 bp in length
 * 29086 29185: gap of unknown length
 * 29186 31257: contig of 2072 bp in length
 * 31258 32673: gap of unknown length
 * 32674 32773: contig of 1316 bp in length
 * 32774 34608: gap of unknown length
 * 34609 34708: gap of unknown length
 * 34709 35916: contig of 1208 bp in length
 * 35917 36016: gap of unknown length
 * 36017 37116: contig of 1100 bp in length
 * 37117 37216: gap of unknown length
 * 37217 39119: contig of 1903 bp in length
 * 39120 39219: gap of unknown length
 * 39220 41424: contig of 2205 bp in length
 * 41425 41524: gap of unknown length
 * 41525 42694: contig of 1170 bp in length
 * 42695 42794: gap of unknown length
 * 42795 44400: contig of 1606 bp in length
 * 44401 44500: gap of unknown length
 * 44501 46996: contig of 2496 bp in length
 * 46997 47096: gap of unknown length
 * 47097 48371: contig of 1275 bp in length
 * 48372 48471: gap of unknown length
 * 48472 50351: contig of 1880 bp in length
 * 50352 50451: gap of unknown length
 * 50452 52100: contig of 1649 bp in length
 * 52101 54048: gap of unknown length
 * 54049 54148: contig of 1848 bp in length
 * 54149 55721: gap of unknown length
 * 55722: contig of 1573 bp in length

* 55722 55821: gap of unknown length
 * 55822 57240: contig of 1419 bp in length
 * 57241 57340: gap of unknown length
 * 57341 60003: contig of 2663 bp in length
 * 60004 60103: gap of unknown length
 * 60104 62111: contig of 2008 bp in length
 * 62112 62211: gap of unknown length
 * 62212 64026: contig of 1815 bp in length
 * 64027 64126: gap of unknown length
 * 64127 66191: contig of 2065 bp in length
 * 66192 66291: gap of unknown length
 * 66292 69826: contig of 3535 bp in length
 * 69827 69926: gap of unknown length
 * 69927 71769: contig of 1843 bp in length
 * 71770 71869: gap of unknown length
 * 71870 73949: contig of 2080 bp in length
 * 73950 74049: gap of unknown length
 * 74050 76696: contig of 2647 bp in length
 * 76697 76796: gap of unknown length
 * 76797 79517: contig of 2721 bp in length
 * 79518 79617: gap of unknown length
 * 79618 81607: contig of 1990 bp in length
 * 81608 81707: gap of unknown length
 * 81708 83237: contig of 1530 bp in length
 * 83238 83337: gap of unknown length
 * 83338 85027: contig of 1690 bp in length
 * 85028 85127: gap of unknown length
 * 85128 88310: contig of 3183 bp in length
 * 88311 88410: gap of unknown length
 * 88411 92204: contig of 3794 bp in length
 * 92205 92304: gap of unknown length
 * 92305 93769: contig of 1465 bp in length
 * 93770 93869: gap of unknown length
 * 93870 96723: contig of 2853 bp in length
 * 96723 96823: gap of unknown length
 * 96823 98907: contig of 2085 bp in length
 * 98908 99007: gap of unknown length
 * 99008 101249: contig of 2242 bp in length
 * 101250 101349: gap of unknown length
 * 101350 103337: contig of 1988 bp in length

Query Match 61.2%; Score 20.8; DB 2; Length 178549;
 Best Local Similarity 78.1%; Pred. No. 97;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 CTCCTACTATGACGCTACGTGAGACA 34
 DB 77988 CTACTATCTATGACAGCGTGTGAGAGA 78019

RESULT 6
 AC009300/c 109939 bp DNA linear PRI 30-SEP-2000
 LOCUS Homo sapiens BAC clone RP11-52A21 from 2, complete sequence.
 DEFINITION AC009300
 ACCESSION AC009300.2 GI:6604538
 VERSION HTG.
 KEYWORDS Homo sapiens.
 SOURCE
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 109939)
 Sultston, J.E. and Waterston, R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 JOURNAL
 MEDLINE 99063792
 PUBMED 9847074
 2 (bases 1 to 109939)
 REFERENCE
 Sun, H., Woldmann, P., Yeakum, M. and Schrader, A.
 The sequence of Homo sapiens BAC clone RP11-52A21
 JOURNAL
 Unpublished
 REFERENCE
 3 (bases 1 to 109939)
 Waterston, R.H.

TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 1099339)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 1099339)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 1099339)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 20, 1999 this sequence version replaced gi:5732174.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
Summary Statistics8
Center project name: H_NH052A21

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,B.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-24K2, 200 bp overlap; the
clone sequenced to the right is RP11-11G16. Actual start of this
clone is at base position 158007 of RP11-24K2.
Location/Qualifiers
1. 109939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-52A21"
/clone_id="RPCI-11"
repeat_region 157..398

/rpc_family="MIR"
repeat_region 1604..2335
/rpc_family="L2"
repeat_region 2616..2722
/rpc_family="L2"
repeat_region 3924..4253
/rpc_family="L2"
repeat_region 4508..5040
/rpc_family="L1"
repeat_region 6995..7105
/rpc_family="L2"
repeat_region 7220..7355
/rpc_family="L2"
repeat_region 7595..7769
/rpc_family="MER1_type"
repeat_region 7970..8235
/rpc_family="Alu"
repeat_region 8550..8595
/rpc_family="AT_rich"
repeat_region 9165..9511
/rpc_family="MALR"
repeat_region 10521..10829
/rpc_family="Alu"
repeat_region 11671..11977
/rpc_family="Alu"
repeat_region 13237..13374
/rpc_family="MER2_type"
repeat_region 13378..13619
/rpc_family="L1"
repeat_region 17112..17346
/rpc_family="MIR"
repeat_region 19007..19311
/rpc_family="Alu"
repeat_region 19661..19939
/rpc_family="Alu"
repeat_region 20109..20181
/rpc_family="MER1_type"
repeat_region 20650..20715
/rpc_family="MER1_type"
repeat_region 20849..20968
/rpc_family="L1"
repeat_region 20969..21344
/rpc_family="MALR"
repeat_region 21345..21969
/rpc_family="MALR"
repeat_region 21970..22338
/rpc_family="MALR"
repeat_region 22465..22712
/rpc_family="L2"
repeat_region 23563..23584
/rpc_family="AT_rich"
repeat_region 24343..24553
/rpc_family="Retroviral"
repeat_region 24971..25144
/rpc_family="Retroviral"
repeat_region 25227..25350
/rpc_family="Retroviral"
repeat_region 26618..26917
/rpc_family="Alu"
repeat_region 26949..27325
/rpc_family="MALR"
repeat_region 27758..27919
/rpc_family="L2"
repeat_region 28384..28451
/rpc_family="L2"
repeat_region 28529..29098
/rpc_family="Retroviral"
repeat_region 29473..29914
/rpc_family="L2"
repeat_region 30014..30038
/rpc_family="(TC)n"
repeat_region 30838..31133
/rpc_family="L1"

```

repeat_region      31405..31699
                    /rpt_family="Alu"
repeat_region      32303..32730
                    /rpt_family="L1"
repeat_region      32755..32829
                    /rpt_family="AT_rich"
repeat_region      32896..32962
                    /rpt_family="MERL_type"
repeat_region      33028..33311
                    /rpt_family="L1"
repeat_region      33481..34158
                    /rpt_family="L1"
repeat_region      34159..35231
                    /rpt_family="retroviral"
repeat_region      35332..35476
                    /rpt_family="L1"
repeat_region      36141..36921
                    /rpt_family="L1"
repeat_region      37104..37147
                    /rpt_family="AT_rich"
repeat_region      38136..38717
                    /rpt_family="L1"
repeat_region      38718..39035
                    /rpt_family="Alu"
repeat_region      39036..39273
                    /rpt_family="L1"
repeat_region      39339..39650
                    /rpt_family="Alu"
repeat_region      40112..40222
                    /rpt_family="L1"
repeat_region      40231..40253
                    /rpt_family="AT_rich"
repeat_region      40440..40635

```

```

Query Match      59.4%; Score 20.2; DB 9; Length 109939;
Best Local Similarity 75.8%; Pred. No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

RESULT 7
LOCUS      AC010249      111081 bp      DNA      linear      HTG 06-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-406F2, WORKING DRAFT SEQUENCE.
ACCESSION  AC010249
VERSION     AC010249.5 GI:7711405
KEYWORDS   HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 11081)
AUTHORS     Makiyoto; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL     Direct Submission
REFERENCE   2 (bases 1 to 11081)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 11081)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On May 6, 2000 this sequence version replaced gi:7710813.
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            Project Information
            Center Project Name: 341206
            Center clone name: CIT-HSPC_406F2

```

```

-----
Summary Statistics
Consensus quality: 104960 bases at least Q40
Consensus quality: 109090 bases at least Q30
Consensus quality: 109623 bases at least Q20
Estimated insert size: 125000; pulse field gel estimation
Estimated insert size: 109981; sum-of-contigs estimation
Quality coverage: 6.74 in Q20 bases; pulse field gel estimation
Quality coverage: 7.62 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
      consists of 12 contigs. Gaps between the contigs
      are represented as runs of N. The order of the pieces
      is believed to be correct as given, however the sizes
      of the gaps between them are based on estimates that have
      been provided by the submitter.
      This sequence will be replaced
      by the finished sequence as soon as it is available and
      the accession number will be preserved.
1 11274: contig of 11274 bp in length
* 11275 11374: gap of unknown length
* 11375 13418: contig of 2044 bp in length
* 13419 13518: gap of unknown length
* 13519 13519: contig of 22821 bp in length
* 13519 36339: gap of unknown length
* 36340 36439: gap of unknown length
* 36440 37882: contig of 1443 bp in length
* 37883 37983: gap of unknown length
* 37983 46574: contig of 8592 bp in length
* 46575 46674: gap of unknown length
* 46675 49885: contig of 3211 bp in length
* 49886 49985: gap of unknown length
* 49986 53530: contig of 3545 bp in length
* 53531 53631: gap of unknown length
* 53631 68020: contig of 14390 bp in length
* 68021 68120: gap of unknown length
* 68121 72250: contig of 4170 bp in length
* 72251 72390: gap of unknown length
* 72391 84023: contig of 11633 bp in length
* 84024 84123: gap of unknown length
* 84124 102743: contig of 18620 bp in length
* 102744 102843: gap of unknown length
* 102844 11081: contig of 8238 bp in length.
Location/Qualifiers
source
1. 111081
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="5"
   /clone="CTC-406F2"
   /clone_lib="Caltech human BAC library C"
BASE COUNT      30700 a 22728 c 24378 g 32141 t 1134 others
ORIGIN

```

```

Query Match      59.4%; Score 20.2; DB 2; Length 111081;
Best Local Similarity 75.8%; Pred. No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

RESULT 8
LOCUS      AC011264/c      137165 bp      DNA      linear      HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-3K8, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC011264
VERSION     AC011264.4 GI:9121012
KEYWORDS   HTG; HTGS PHASE0.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 137165)
AUTHORS     Makiyoto; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Birren,B., Linton,L., Nuebaum,C. and Lander,B.

```

TITLE Homo sapiens, clone RP11-3K8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137165)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguski, L., Bouckgeat, B., Brown, A., Castle, A., Collins, S., Collymore, A., Cooke, P., D'Amico, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardina, S., Grant, G., Haggas, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karst, A., Klein, J., Lechoczy, J., Liu, C., Locke, K., Macdonald, P., Marquis, N., McEwen, P., McQuirk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L2766
Center clone name: 3_K_8

NOTE: This record contains 161 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 650: contig of 650 bp in length
* 651 750: gap of 100 bp
* 751 1461: contig of 711 bp in length
* 1462 1561: gap of 100 bp
* 1562 2270: contig of 709 bp in length
* 2271 2370: gap of 100 bp
* 2371 3065: contig of 695 bp in length
* 3066 3165: gap of 100 bp
* 3166 3782: contig of 617 bp in length
* 3783 3882: gap of 100 bp
* 3883 4535: contig of 653 bp in length
* 4536 4635: gap of 100 bp
* 4636 5316: contig of 681 bp in length
* 5317 5416: gap of 100 bp
* 5417 6092: contig of 676 bp in length
* 6093 6192: gap of 100 bp
* 6193 6875: contig of 683 bp in length
* 6876 6975: gap of 100 bp
* 6976 7662: contig of 687 bp in length
* 7663 7762: gap of 100 bp
* 7763 8453: contig of 691 bp in length
* 8454 8553: gap of 100 bp
* 8554 9244: contig of 691 bp in length
* 9245 9344: gap of 100 bp
* 9345 10001: contig of 657 bp in length
* 10002 10101: gap of 100 bp
* 10102 10795: contig of 694 bp in length
* 10796 10895: gap of 100 bp
* 10896 11595: contig of 700 bp in length

11596 11695: gap of 100 bp
* 11696 12396: contig of 701 bp in length
* 12397 12496: gap of 100 bp
* 12497 13154: contig of 658 bp in length
* 13155 13254: gap of 100 bp
* 13255 13924: contig of 670 bp in length
* 13925 14024: gap of 100 bp
* 14025 14712: contig of 688 bp in length
* 14713 14812: gap of 100 bp
* 14813 15486: contig of 674 bp in length
* 15487 15586: gap of 100 bp
* 15587 16278: contig of 692 bp in length
* 16279 16378: gap of 100 bp
* 16379 17064: contig of 686 bp in length
* 17065 17164: gap of 100 bp
* 17165 17872: contig of 708 bp in length
* 17873 17972: gap of 100 bp
* 17973 18636: contig of 664 bp in length
* 18637 18736: gap of 100 bp
* 18737 19451: contig of 715 bp in length
* 19452 19551: gap of 100 bp
* 19552 20257: contig of 706 bp in length
* 20258 20357: gap of 100 bp
* 20358 21037: contig of 680 bp in length
* 21038 21137: gap of 100 bp
* 21138 21776: contig of 639 bp in length
* 21777 21876: gap of 100 bp
* 21877 22518: contig of 642 bp in length
* 22519 22618: gap of 100 bp
* 22619 23263: contig of 645 bp in length
* 23264 23363: gap of 100 bp
* 23364 24044: contig of 681 bp in length
* 24045 24144: gap of 100 bp
* 24145 24800: contig of 656 bp in length
* 24801 24900: gap of 100 bp
* 24901 25563: contig of 663 bp in length
* 25564 25663: gap of 100 bp
* 25664 26345: contig of 682 bp in length
* 26346 26445: gap of 100 bp
* 26446 27145: contig of 700 bp in length
* 27146 27245: gap of 100 bp
* 27246 27904: contig of 659 bp in length
* 27905 28004: gap of 100 bp
* 28005 28713: contig of 709 bp in length
* 28714 28813: gap of 100 bp
* 28814 29513: contig of 700 bp in length
* 29514 29613: gap of 100 bp
* 29614 30322: contig of 709 bp in length
* 30323 30422: gap of 100 bp
* 30423 31078: contig of 656 bp in length
* 31079 31178: gap of 100 bp
* 31179 31829: contig of 651 bp in length
* 31830 31929: gap of 100 bp
* 31930 32579: contig of 650 bp in length
* 32580 32679: gap of 100 bp
* 32680 33364: contig of 685 bp in length
* 33365 33464: gap of 100 bp
* 33465 34150: contig of 686 bp in length
* 34151 34250: gap of 100 bp
* 34251 34917: contig of 667 bp in length
* 34918 35017: gap of 100 bp
* 35018 35721: contig of 704 bp in length
* 35722 35821: gap of 100 bp
* 35822 36538: contig of 717 bp in length
* 36539 36638: gap of 100 bp
* 36639 37301: contig of 663 bp in length
* 37302 37401: gap of 100 bp
* 37402 38117: contig of 716 bp in length
* 38118 38217: gap of 100 bp
* 38218 38912: contig of 695 bp in length
* 38913 39012: gap of 100 bp
* 39013 39715: contig of 703 bp in length
* 39716 39815: gap of 100 bp

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Peters, L., Pickens, R., Primus, E., Pu, L. M., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swalek, A., Taber, P., Tameria, A., Tameria, K., Tang, H.,
 Tamey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, C.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 176172)
 Worley, K. C.
 Direct Submission
 Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 176172)
 Worley, K. C.
 Direct Submission
 Submitted (18-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20806189.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GVRG
 Center clone name: CH230-129CS
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 115261 bases at least Q40
 Consensus quality: 121020 bases at least Q30
 Consensus quality: 124933 bases at least Q20

 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 70 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	1118:	contig of 1118 bp in length	14844	14943:	gap of unknown length
1119	1218:	gap of unknown length	14944	15236:	contig of 1293 bp in length
1219	2651:	contig of 1433 bp in length	16237	16336:	gap of unknown length
2652	2751:	gap of unknown length	16337	17557:	contig of 1221 bp in length
2752	4118:	contig of 1367 bp in length	17558	17657:	gap of unknown length
4119	4218:	gap of unknown length	17658	18680:	contig of 1023 bp in length
4219	5515:	contig of 1297 bp in length	18681	18780:	gap of unknown length
5516	5615:	gap of unknown length	18781	19823:	contig of 1043 bp in length
5616	6879:	contig of 1264 bp in length	19824	19923:	gap of unknown length
6880	6979:	gap of unknown length	19924	21040:	contig of 1117 bp in length
6980	8084:	contig of 1105 bp in length	21041	21140:	gap of unknown length
8085	8184:	gap of unknown length	21141	22332:	contig of 1192 bp in length
8185	9269:	contig of 1085 bp in length	22333	22433:	gap of unknown length
9270	9369:	gap of unknown length	22433	22439:	contig of 1007 bp in length
9370	10868:	contig of 1499 bp in length	22440	23539:	gap of unknown length
10869	10968:	gap of unknown length	23540	24639:	contig of 1100 bp in length
10969	12116:	contig of 1148 bp in length	24640	24739:	gap of unknown length
12117	12216:	gap of unknown length	24740	26190:	contig of 1451 bp in length
12217	13432:	contig of 1216 bp in length	26191	26290:	gap of unknown length
13433	13532:	gap of unknown length	26291	27930:	contig of 1640 bp in length
13533	14843:	contig of 1311 bp in length	27931	28030:	gap of unknown length
			28031	29345:	contig of 1315 bp in length
			29346	29445:	gap of unknown length
			29446	30567:	contig of 1122 bp in length
			30568	30667:	gap of unknown length
			30668	32268:	contig of 1601 bp in length
			32269	32368:	gap of unknown length
			32369	33885:	contig of 1517 bp in length
			33886	33985:	gap of unknown length
			33986	35460:	contig of 1475 bp in length
			35461	35560:	gap of unknown length
			35561	37290:	contig of 1729 bp in length
			37290	37389:	gap of unknown length
			37389	39446:	contig of 2057 bp in length
			39447	39546:	gap of unknown length
			39547	41715:	contig of 2169 bp in length
			41716	41815:	gap of unknown length
			41816	43126:	contig of 1311 bp in length
			43127	43266:	gap of unknown length
			43267	44501:	contig of 1275 bp in length
			44502	46501:	gap of unknown length
			46502	46557:	contig of 1656 bp in length
			46558	46557:	gap of unknown length
			46558	48457:	contig of 2100 bp in length
			48458	48557:	gap of unknown length

* 82415: contig of 3217 bp in length
* 85632: gap of unknown length
* 85731: contig of 2161 bp in length
* 87892: gap of unknown length
* 87992: gap of unknown length
* 87993: contig of 1871 bp in length
* 89863: gap of unknown length
* 89864: gap of unknown length
* 89964: contig of 3620 bp in length
* 93584: gap of unknown length
* 93683: contig of 2566 bp in length
* 96250: gap of unknown length
* 96349: gap of 2407 bp in length
* 96350: contig of 2407 bp in length
* 98757: gap of unknown length
* 98856: gap of unknown length
* 98857: contig of 1027 bp in length

Query Match
Best Local Similarity 75.8%; Pred. No. 1.8e+02; Length 176172;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTATCGATCGTACTGTAGAC 33
Db 122052 GCTCCTCTACACGATCGTACTGTAGAC 122020

RESULT 11
AC022438/c
LOCUS
DEFINITION
AC022438 Homo sapiens chromosome 5 clone RP11-137P5, WORKING DRAFT SEQUENCE,
9 ordered pieces.
AC022438 AC022438.4 GI:14579711
VERSION
HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
Homo sapiens.
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 185791)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 185791)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 1, 2001 this sequence version replaced gi:7711740.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 459338
Center clone name: RP11-137P5

Summary Statistics
Consensus quality: 181059 bases at least Q40
Consensus quality: 183500 bases at least Q30
Consensus quality: 184433 bases at least Q20
Estimated insert size: 198400; agarose-ef estimation
Estimated insert size: 184991; sum-of-contigs estimation
Quality coverage: 7.42 in Q20 bases; agarose-ef estimation
Quality coverage: 7.96 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 71478: contig of 71478 bp in length
* 71479 71578: gap of unknown length

* 71579 80602: contig of 9024 bp in length
* 80603 80702: gap of unknown length
* 80703 93617: contig of 12915 bp in length
* 93618 93717: gap of unknown length
* 93718 116162: gap of 22445 bp in length
* 116163 116262: gap of unknown length
* 116263 118116: contig of 1854 bp in length
* 118117 118216: gap of unknown length
* 118217 137813: contig of 19597 bp in length
* 137814 137913: gap of unknown length
* 137914 139655: contig of 1742 bp in length
* 139656 139755: gap of unknown length
* 139756 177952: contig of 38197 bp in length
* 177953 178052: gap of unknown length
* 178053 185791: contig of 7739 bp in length.

FEATURES
Source
1. 185791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-137P5"
/clone_1ib="RP11-137P5"
Location/Qualifiers

BASE COUNT 53841 a 39773 c 38649 g 52728 t 800 others
ORIGIN

Query Match
Best Local Similarity 75.8%; Pred. No. 1.8e+02; Length 185791;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ACTCCTCTACTATCGATCGTACTGTAGACA 34
Db 66938 ACTGCTCACATATCGATCGTCTTAGAACA 66906

RESULT 12
AC130200/c
LOCUS
DEFINITION
AC130200 Medicago truncatula clone mchl-64n13, WORKING DRAFT SEQUENCE, 28
unordered pieces.
AC130200 AC130200.5 GI:22297341
VERSION
HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
barrel medic.
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 80229)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
Medicago truncatula BAC Clone mchl-64n13
Unpublished
2 (bases 1 to 80229)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
Direct Submission
Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 80229)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
Direct Submission
Submitted (19-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Aug 19, 2002 this sequence version replaced gi:22218474.
-----Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 1999: contig of 1999 bp in length
* 2000 2099: gap of unknown length
* 2100 4226: contig of 2127 bp in length
* 4227 4326: gap of unknown length
* 4327 6630: contig of 2304 bp in length
* 6631 6730: gap of unknown length
* 6731 9394: contig of 2664 bp in length
* 9395 9494: gap of unknown length
* 9495 12005: contig of 2511 bp in length
* 12006 12105: gap of unknown length
* 12106 14839: contig of 2734 bp in length
* 14840 14939: gap of unknown length
* 14940 17314: contig of 2375 bp in length
* 17315 17414: gap of unknown length
* 17415 19640: contig of 2226 bp in length
* 19641 19740: gap of unknown length
* 19741 21813: contig of 2073 bp in length
* 21814 25010: gap of unknown length
* 25011 25110: contig of 3097 bp in length
* 25111 28032: contig of 2922 bp in length
* 28033 28132: gap of unknown length
* 30281 30282: contig of 2149 bp in length
* 30283 30381: gap of unknown length
* 30382 33040: contig of 2659 bp in length
* 33041 33140: gap of unknown length
* 33141 35458: contig of 2318 bp in length
* 35459 35558: gap of unknown length
* 35559 37727: contig of 2169 bp in length
* 37728 37827: gap of unknown length
* 37828 39889: contig of 2062 bp in length
* 39890 39989: gap of unknown length
* 39990 42308: contig of 2319 bp in length
* 42309 42408: gap of unknown length
* 42409 44708: contig of 2300 bp in length
* 44709 44808: gap of unknown length
* 44809 47075: contig of 2267 bp in length
* 47076 47175: gap of unknown length
* 47176 50293: contig of 3118 bp in length
* 50294 50393: gap of unknown length
* 50394 52798: contig of 2405 bp in length
* 52799 52898: gap of unknown length
* 52899 57000: contig of 4102 bp in length
* 57001 57100: gap of unknown length
* 57101 60039: contig of 2939 bp in length
* 60040 60139: gap of unknown length
* 60140 63242: contig of 3103 bp in length
* 63243 63342: gap of unknown length
* 63343 66407: contig of 3065 bp in length
* 66408 66507: gap of unknown length
* 66508 72073: contig of 5566 bp in length
* 72074 72173: gap of unknown length
* 72174 77199: contig of 5026 bp in length
* 77200 77299: gap of unknown length
* 77300 80229: contig of 2930 bp in length.
```

FEATURES

source

1. 80229
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_11b="Medicago truncatula BAC library H1"

BASE COUNT 25729 a 13595 c 12936 g 25254 t 2715 others

Query Match

58.8%; Score 20; DB 2; Length 80229;

Best Local Similarity 82.1%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCTCCACTTATCGATCGTACTG 29
Db 43490 ACTCTCACTTTGATCGTACTG 43463

RESULT 13

AC121899/c

LOCUS AC121899 159746 bp DNA linear HTG 21-MAY-2002
DEFINITION Mus musculus chromosome UNK clone RP24-166J12, WORKING DRAFT
SEQUENCE, 4 unordered pieces.

AC121899

AC121899

AC121899.1

GI:21040020

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

McPherson, J.D. and Waterston, R.H.

1 (bases 1 to 159746)

2 (bases 1 to 159746)

Unpublished

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT

Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

Contact: submissions@wustl.wustl.edu

Project Information

Center project name: M B80166J12

Summary Statistics

Sequencing vector: MJ3, 0%

Sequencing vector: plasmid, 100%

Chemistry: Dye-primer ET; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 15759 bases at least Q40

Consensus quality: 15836 bases at least Q30

Insert size: 159400; agarose-fp

Quality coverage: 16.21 in Q20 bases; agarose-fp

Quality coverage: 12.58 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

consists of 4 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 8022

8121: gap of unknown length

8122 26511: gap of 18390 bp in length

26512 26611: gap of unknown length

26612 76565: contig of 49954 bp in length

76566 76665: gap of unknown length

159746: contig of 83081 bp in length.

Location/Qualifiers

1. 159746

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="UNK"

/clone="RP24-166J12"

```

misc_feature      1. .8021
                  /note="assembly_name:Contig12"
misc_feature      8122. .26511
                  /note="assembly_name:Contig13"
misc_feature      26612. .76565
                  /note="assembly_name:Contig14"
misc_feature      76666. .159746
                  /note="assembly_name:Contig15"
BASE COUNT      51030 a 29319 c 30906 g 48183 t 308 others
ORIGIN
Query Match      58.84; Score 20; DB 2; Length 159746;
Best Local Similarity 82.14; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 7 TCTATTATGATGATGCTGCTGAGACA 34
Db 81057 TCTATTATGCTCTGTCTTGAGACA 81030

```

```

RESULT 14
LOCUS AC107106 170083 bp DNA linear HTG 17-JUL-2002
DEFINITION Rattus norvegicus clone CH230-155M2. *** SEQUENCING IN PROGRESS
ACCESSION AC107106.3 GI:21743917
VERSION HTG; HTGS PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 170083)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooke,S.L., Amaralung,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homesi,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulesged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Meheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Monabadi,K., Morgan,M., Morris,L.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogun,M., Okumotu,G.,
Otaguave,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Prime,E., Pu,L.L., Quilez,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Schierer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Tabor,P., Tameria,A., Tamerias,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tumait,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,

```

```

TITLE      Weinstock,G. and Gibbs,R.
JOURNAL    Direct Submission
REFERENCE  2 (bases 1 to 170083)
AUTHORS    Morley,K.C.
TITL       Direct Submission
JOURNAL    Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 170083)
AUTHORS    Morley,K.C.
TITL       Direct Submission
JOURNAL    Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:18701531.
COMMENT    ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNIF
Center clone name: CH230-155M2
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990129
Consensus quality: 111369 bases at least Q40
Consensus quality: 116447 bases at least Q30
Consensus quality: 121122 bases at least Q20
-----

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1039: contig of 1039 bp in length
* 1040 1139: gap of unknown length
* 1140 2259: contig of 1120 bp in length
* 2260 2359: gap of unknown length
* 2360 3500: contig of 1141 bp in length
* 3501 3600: gap of unknown length
* 3601 4883: contig of 1283 bp in length
* 4884 4983: gap of unknown length
* 4984 6335: contig of 1352 bp in length
* 6336 6435: gap of unknown length
* 6436 7741: contig of 1306 bp in length
* 7742 7841: gap of unknown length
* 7842 9151: contig of 1310 bp in length
* 9152 9251: gap of unknown length
* 9252 10753: contig of 1502 bp in length
* 10754 10853: gap of unknown length
* 10854 12543: contig of 1690 bp in length
* 12544 12643: gap of unknown length
* 12644 13865: contig of 1222 bp in length
* 13866 13965: gap of unknown length
* 13966 15202: contig of 1231 bp in length
* 15203 15302: gap of unknown length
* 15303 16317: contig of 1015 bp in length
* 16318 16417: gap of unknown length
* 16418 17872: contig of 1455 bp in length
* 17873 17972: gap of unknown length
* 17973 19199: contig of 1227 bp in length
* 19200 19299: gap of unknown length
* 19300 20563: contig of 1264 bp in length
* 20564 22126: gap of unknown length
* 22126: contig of 1463 bp in length

```

```

* 22127 22226: gap of unknown length
* 22227 23739: contig of 1513 bp in length
* 23740 23839: gap of unknown length
* 23840 25334: contig of 1495 bp in length
* 25335 25434: gap of unknown length
* 25435 27296: contig of 1862 bp in length
* 27297 27396: gap of unknown length
* 28964 29063: gap of unknown length
* 29064 30623: contig of 1559 bp in length
* 30623 30723: gap of unknown length
* 30723 31938: contig of 1217 bp in length
* 31940 32039: gap of unknown length
* 32040 34250: contig of 2211 bp in length
* 34251 34350: gap of unknown length
* 34351 36580: contig of 2230 bp in length
* 36581 36680: gap of unknown length
* 38029: contig of 1349 bp in length
* 38030 38129: gap of unknown length
* 38130 40235: contig of 2106 bp in length
* 40236 40335: gap of unknown length
* 40336 42225: contig of 1890 bp in length
* 42226 42325: gap of unknown length
* 42326 44117: contig of 1792 bp in length
* 44118 44217: gap of unknown length
* 44218 45978: contig of 1761 bp in length
* 45979 46078: gap of unknown length
* 46079 47659: contig of 1581 bp in length
* 47660 49260: gap of unknown length
* 49261 49360: gap of unknown length
* 49361 51125: contig of 1765 bp in length
* 51126 51225: gap of unknown length
* 51226 54076: contig of 2851 bp in length
* 54077 54176: gap of unknown length
* 54177 55760: contig of 1584 bp in length
* 55761 55860: gap of unknown length
* 55861 58528: contig of 2668 bp in length
* 58529 58628: gap of unknown length
* 58629 61885: contig of 3257 bp in length
* 61886 61985: gap of unknown length
* 61986 65195: contig of 3210 bp in length
* 65196 65295: gap of unknown length
* 65296 67635: contig of 2340 bp in length
* 67636 67735: gap of unknown length
* 67736 70102: contig of 2367 bp in length
* 70103 70202: gap of unknown length
* 70203 73728: contig of 3526 bp in length
* 73729 73828: gap of unknown length
* 73829 76882: contig of 3054 bp in length
* 76883 76982: gap of unknown length
* 76983 80139: contig of 3157 bp in length
* 80140 80239: gap of unknown length
* 80240 82873: gap of unknown length
* 82874 82973: gap of unknown length
* 82974 85025: contig of 2052 bp in length
* 85026 85125: gap of unknown length
* 85126 88495: contig of 4370 bp in length
* 88496 88595: gap of unknown length
* 88596 92336: contig of 2741 bp in length
* 92337 92436: gap of unknown length
* 92437 95301: contig of 2865 bp in length
* 95302 95401: gap of unknown length
* 95402 100512: contig of 5111 bp in length
* 100513 100612: gap of unknown length
* 100613 103557: contig of 2945 bp in length
* 103558 103657: gap of unknown length
* 103658 107805: contig of 4148 bp in length
* 107806 107905: gap of unknown length
* 107906 110962: contig of 3057 bp in length
* 110963 111062: gap of unknown length
* 111063 114462: contig of 3400 bp in length
* 114463 114562: gap of unknown length

```

```

* 114563 117995: contig of 3433 bp in length
* 117996 118095: gap of unknown length
* 118096 121626: contig of 3531 bp in length

Query Match      58.8%; Score 20; DB 2; Length 170083;
Best Local Similarity 82.1%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      4  TCCCTACTATCATCGCTACTGTGAG 31
Db 60176 TCCATCTACTATCATGATGATCTGGAAG 60203

```

```

RESULT 15
AC129318/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 227789)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
2 (bases 1 to 227789)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watsn.wustl.edu
Center project name: M_BA0464B17
----- Project Information -----

```

```

----- Summary Statistics -----
Sequencing vector: MJ3; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 222687 bases at least Q40
Consensus quality: 223874 bases at least Q30
Consensus quality: 224621 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 226459; sum-of-contigs
Quality coverage: 17.88 in Q20 bases; agarose-fp
Quality coverage: 11.92 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1364: contig of 1364 bp in length
* 1365 1464: gap of unknown length
* 1465 3160: contig of 1696 bp in length
* 3161 3260: gap of unknown length
* 3261 5330: contig of 2060 bp in length
* 5331 5420: gap of unknown length
* 5421 6609: contig of 1189 bp in length
* 6610 6709: gap of unknown length

```

FEATURES	
*	6710
*	24323: contig of 17614 bp in length
*	24324
*	24423: gap of unknown length
*	24454
*	31816: contig of 7393 bp in length
*	31817
*	31916: gap of unknown length
*	31917
*	42488: contig of 10572 bp in length
*	42489
*	42588: gap of unknown length
*	42589
*	42718: contig of 130 bp in length
*	42719
*	42818: gap of unknown length
*	42819
*	73139: contig of 30321 bp in length
*	73140
*	73239: gap of unknown length
*	73240
*	97703: contig of 24464 bp in length
*	97704
*	97803: gap of unknown length
*	97804
*	137021: contig of 39218 bp in length
*	137022
*	180567: contig of 43446 bp in length
*	137712
*	180667: gap of unknown length
*	180668
*	227789: contig of 47122 bp in length
	Location/Qualifiers

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-464B17"
1..1364
/note="assembly_name:Cont1g49"
1465..3160
/note="assembly_name:Cont1g51"
3261..5320
/note="assembly_name:Cont1g52"
5421..6609
/note="assembly_name:Cont1g53"
6710..24323
/note="assembly_name:Cont1g54"
24424..31816
/note="assembly_name:Cont1g55"
31917..42488
/note="assembly_name:Cont1g56"
42589..42718
/note="assembly_name:Cont1g511"
42819..773139
/note="assembly_name:Cont1g57"
73240..97703
/note="assembly_name:Cont1g58"
97804..137021
/note="assembly_name:Cont1g59"
137122..180567
/note="assembly_name:Cont1g60"
180668..222789
/note="assembly_name:Cont1g61"
40836 c 14553 g 73395 t 1217 others

```

Query Match	58.8%	Score 20	Length 227789
Best Local Similarity	82.1%	Pred. No. 2.2e+02	
Matches 23	Conservative 0	Mismatches 5	Indels 0
Gaps 0			
QY	7	TCTACTATCGATCGCTACTGTGAACA	34
DB	11101	TCTACTATCTCTCGTCTCTTGAACA	11074

Search completed: May 11, 2003, 06:08:20
Job time : 558.11 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1269.94 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgctgcagactagtgatt.....aaggatgcagcatgccaact 79

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: em.aethba.*
- 2: em.aethba.*
- 3: em.aethin.*
- 4: em.aethm.*
- 5: em.aethov.*
- 6: em.aethpl.*
- 7: em.aethro.*
- 8: em.aethrc.*
- 9: gb.aethrc.*
- 10: gb.aethrc.*
- 11: gb.aethrc.*
- 12: gb.aethrc.*
- 13: gb.aethrc.*
- 14: gb.aethrc.*
- 15: em.aethrc.*
- 16: em.aethom.*
- 17: gb.aethrc.*
- 18: em.aethrc.*
- 19: em.aethrc.*
- 20: em.aethrc.*
- 21: em.aethrc.*
- 22: em.aethrc.*
- 23: em.aethrc.*
- 24: em.aethrc.*
- 25: em.aethrc.*
- 26: em.aethrc.*
- 27: em.aethrc.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	43	54.4	224	17	BH746858 SALK_0036
2	39	49.4	112	17	BH751161 SALK_0496
3	39	49.4	112	17	BH749349 SALK_0477
4	39	49.4	153	17	BH619283 SALK_0407
5	39	49.4	153	17	BH747013 SALK_0080
6	39	49.4	158	17	BH748291 SALK_0451

7	39	165	17	BH747357	SALK	0165	BH747357	SALK	0165	
8	39	49.4	175	17	BH746616	SALK	0452	BH746616	SALK	0452
C 9	39	49.4	177	17	BH802465	1008026E1	BH802465	1008026E1	BH802465	1008026E1
10	39	49.4	190	17	BH752801	SALK	0193	BH752801	SALK	0193
11	39	49.4	193	17	BH748289	SALK	0450	BH748289	SALK	0450
12	39	49.4	214	17	BH747839	SALK	0353	BH747839	SALK	0353
13	39	49.4	215	17	BH753813	SALK	0296	BH753813	SALK	0296
14	39	49.4	219	17	BH747744	SALK	0337	BH747744	SALK	0337
15	39	49.4	220	17	BH747438	SALK	0174	BH747438	SALK	0174
C 16	39	49.4	221	17	BH802415	1008026B0	BH802415	1008026B0	BH802415	1008026B0
17	39	49.4	230	17	BH746474	SALK	0428	BH746474	SALK	0428
C 18	39	49.4	230	17	BH799173	1008025B1	BH799173	1008025B1	BH799173	1008025B1
C 19	39	49.4	237	17	BH802463	1008026E0	BH802463	1008026E0	BH802463	1008026E0
C 20	39	49.4	244	17	BH746375	SALK	0402	BH746375	SALK	0402
C 21	39	49.4	248	17	BH254738	SALK	0172	BH254738	SALK	0172
22	39	49.4	251	17	BH750171	SALK	0371	BH750171	SALK	0371
23	39	49.4	254	17	BH748500	SALK	0460	BH748500	SALK	0460
24	39	49.4	256	17	BH748459	SALK	0460	BH748459	SALK	0460
C 25	39	49.4	256	17	BH802470	1008026F0	BH802470	1008026F0	BH802470	1008026F0
C 26	39	49.4	261	17	BH802493	1008026H0	BH802493	1008026H0	BH802493	1008026H0
C 27	39	49.4	268	17	BH611919	SALK	0318	BH611919	SALK	0318
C 28	39	49.4	269	17	BH211646	SALK	0064	BH211646	SALK	0064
C 29	39	49.4	269	17	BH802489	1008026G0	BH802489	1008026G0	BH802489	1008026G0
C 30	39	49.4	271	17	BH799178	1008025C0	BH799178	1008025C0	BH799178	1008025C0
C 31	39	49.4	272	17	BH802438	1008026C0	BH802438	1008026C0	BH802438	1008026C0
C 32	39	49.4	274	17	BH802485	1008026H0	BH802485	1008026H0	BH802485	1008026H0
C 33	39	49.4	279	17	BH748475	SALK	0460	BH748475	SALK	0460
C 34	39	49.4	281	17	BH750170	SALK	0371	BH750170	SALK	0371
C 35	39	49.4	281	17	BH802443	1008026D0	BH802443	1008026D0	BH802443	1008026D0
C 36	39	49.4	284	17	BH213307	SALK	0090	BH213307	SALK	0090
37	39	49.4	284	17	BH746665	SALK	0457	BH746665	SALK	0457
38	39	49.4	286	17	BH747007	SALK	0078	BH747007	SALK	0078
C 39	39	49.4	293	17	BH610310	SALK	0085	BH610310	SALK	0085
C 40	39	49.4	296	17	BH213532	SALK	0093	BH213532	SALK	0093
C 41	39	49.4	335	17	BH213230	SALK	0089	BH213230	SALK	0089
C 42	39	49.4	344	17	BH747662	SALK	0323	BH747662	SALK	0323
43	39	49.4	350	17	BH747675	SALK	0325	BH747675	SALK	0325
44	39	49.4	355	17	BH747371	SALK	0166	BH747371	SALK	0166
45	39	49.4	360	17	BH747123	SALK	0107	BH747123	SALK	0107

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION
BH746858	224 bp · DNA	linear GSS 27-FEB-2002
BH746858	SAUK_003694.51.40.x	Arabidopsis thaliana TDNA insertion lines
	Arabidopsis thaliana genomic clone SAUK_003694.51.40.x	DNA

ACCESSION	BH746858
VERSION	BH746858.1
	GI:18959973

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

TITLE	A Sequence-Indexed Library of Insertion Mutations in the
JOURNAL	Arabidopsis Genome
COMMENT	Unpublished (2001)
	Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)
Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckerk@salk.edu

This is single pass sequence recovered from the left

DEFINITION SALK_040762 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_040762, DNA sequence.

ACCESSION BH619283

VERSION BH619283.1 GI:18429799

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 153)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..153
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_040762"
/note="11b="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 36 a 34 c 36 g 47 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
DB 110 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72

RESULT 5
BH747013 153 bp DNA linear GSS 27-FEB-2002

LOCUS SALK_008070.43.05.x Arabidopsis thaliana TDNA insertion lines

DEFINITION Arabidopsis thaliana genomic clone SALK_008070.43.05.x, DNA sequence.

ACCESSION BH747013

VERSION BH747013.1 GI:18960128

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 153)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..158
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_045100.51.10.x"
/clone_11b="Arabidopsis thaliana TDNA insertion lines"

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..153
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_008070.43.05.x"
/clone_11b="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 47 a 37 c 33 g 36 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
DB 44 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 82

RESULT 6
BH748291 158 bp DNA linear GSS 27-FEB-2002

LOCUS SALK_045100.51.10.x Arabidopsis thaliana TDNA insertion lines

DEFINITION Arabidopsis thaliana genomic clone SALK_045100.51.10.x, DNA sequence.

ACCESSION BH748291

VERSION BH748291.1 GI:18961648

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 158)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..158
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_045100.51.10.x"
/clone_11b="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT

37 a 35 c 37 g 49 t

Query Match

Best Local Similarity 49.4%; Score 39; DB 17; Length 158; Pred. No. 0.0036; Mismatches 0; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7

BH747357

LOCUS 165 bp DNA linear GSS 27-FEB-2002

DEFINITION SALK_016522.55.50.x Arabidopsis thaliana TDNA insertion lines

Arabidopsis thaliana genomic clone SALK_016522.55.50.x, DNA

sequence.

ACCESSION BH747357

VERSION BH747357.1

GI:18960472

KEYWORDS

GSS.

SOURCE

Thale cress.

ORGANISM

Arabidopsis thaliana

REFERENCE

1

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab,

TITLE

A Sequence-Indexed Library of Insertion Mutations in the

JOURNAL

Arabidopsis Genome

COMMENT

Unpublished (2001)

CONTACT

Contact: Joseph R. Ecker

SALK INSTITUTE

Salk Institute Genomic Analysis Laboratory (SIGNAL)

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

Class: TDNA tagged.

Location/Qualifiers

1. 165

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_016522.55.50.x"

/notes="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT

45 a 43 c 39 g 38 t

Query Match

Best Local Similarity 49.4%; Score 39; DB 17; Length 165;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

Db 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

Db 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

RESULT 8

BH746616 175 bp DNA linear GSS 27-FEB-2002

LOCUS

SALK_045268.54.50.x Arabidopsis thaliana TDNA insertion lines

DEFINITION

Arabidopsis thaliana genomic clone SALK_045268.54.50.x, DNA

sequence.

ACCESSION

BH746616

VERSION

BH746616.1

KEYWORDS

GSS.

SOURCE

Thale cress.

ORGANISM

Arabidopsis thaliana

REFERENCE

1

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab,

TITLE

A Sequence-Indexed Library of Insertion Mutations in the

JOURNAL

Arabidopsis Genome

COMMENT

Unpublished (2001)

CONTACT

Contact: Joseph R. Ecker

SALK INSTITUTE

Salk Institute Genomic Analysis Laboratory (SIGNAL)

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

Class: TDNA tagged.

Location/Qualifiers

1. 175

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_045268.54.50.x"

/notes="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT

54 a 43 c 40 g 38 t

Query Match

Best Local Similarity 49.4%; Score 39; DB 17; Length 175;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

Db 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

RESULT 9

BH802465

LOCUS

BH802465

DEFINITION

1008026E10.y2 1008 - Rescuedu Grid 1 Zea mays genomic, DNA

sequence.

BH802465

ACCESSION

BH802465

VERSION

BH802465.1

KEYWORDS

GSS.

SOURCE

Zea mays.

ORGANISM

Zea mays

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 177)

Walbot,V.

TITLE

Maize genomic sequences found using engineered Rescuedu transposon

JOURNAL COMMENT

Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Very probable ligation site found so sequence was trimmed.
Post-ligation sequence submitted separately.
Plate: 1008026 row: 10
Class: transposon-tagged.

FEATURES

source

Location/Qualifiers
1. 177
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/clone_id="1008 - RescueMu Grid 1"
/issue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: leaf; Vector: RescueMu (engineered from pBluescript backbone); Site 1: BamHI, Site 2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site www.zmbl.iasate.edu and follow the links for 'RescueMu.' Grid 1 was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT

39 a 38 c 47 g 53 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.0038;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAGGAGTACGCA 72
|||||
DB 118 GATTGATGTGATATCTCCACTGACGTAGGAGTACGCA 80
|||||

RESULT 10

LOCUS

BH752801 190 bp DNA linear GSS 27-FEB-2002

DEFINITION

SALK_019366.54.25.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_019366.54.25.x, DNA
sequence.

ACCESSION

BH752801

VERSION

BH752801.1 GI:18972023

KEYWORDS

GSS.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadri nab
1 (bases 1 to 190)
, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
, Zimmerman,J. and Ecker,J.R.

TITLE

A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome

JOURNAL

Unpublished (2001)
Contact: Joseph R. Ecker

The Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

Email: ecker@salk.edu

FEATURES

source

Location/Qualifiers
1. 190
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_id="SALK_019366.54.25.x"

/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT

52 a 52 c 41 g 45 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAGGAGTACGCA 72
|||||
DB 81 GATTGATGTGATATCTCCACTGACGTAGGAGTACGCA 119
|||||

RESULT 11

LOCUS

BH748289 193 bp DNA linear GSS 27-FEB-2002

DEFINITION

SALK_045097.54.25.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_045097.54.25.x, DNA
sequence.

ACCESSION

BH748289
BH748289.1 GI:18961643

VERSION

GSS.

KEYWORDS

thale cress.

SOURCE

Arabidopsis thaliana

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadri nab
1 (bases 1 to 193)
, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
, Zimmerman,J. and Ecker,J.R.

TITLE

A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome

JOURNAL

Unpublished (2001)
Contact: Joseph R. Ecker

The Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

REFERENCE

AUTHORS

Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

FEATURES

source

Location/Qualifiers
1. 193
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"

/clone_id="SALK_045097.54.25.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT

61 a 48 c 43 g 41 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
DB 84 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 122
|||||

RESULT 12

BH747829 214 bp DNA linear GSS 27-FEB-2002
LOCUS SALK_035354.55.75.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_035354.55.75.x, DNA
sequence.

ACCESSION BH747829
VERSION BH747829
KEYWORDS GS8.
SOURCE GI:18960944
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 214)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..214
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_035354.55.75.x"
/note="lib="Arabidopsis thaliana TDNA insertion lines"
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 58 a 56 c 52 g 48 t

Query Match 49.4%; Score 39; DB 17; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
DB 104 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 142
|||||

RESULT 13

BH753813 215 bp DNA linear GSS 27-FEB-2002
LOCUS SALK_029677.55.00.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_029677.55.00.x, DNA
sequence.

ACCESSION BH753813
VERSION BH753813
KEYWORDS GS8.
SOURCE GI:18974308
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 215)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..215
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_029677.55.00.x"
/note="lib="Arabidopsis thaliana TDNA insertion lines"
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 63 a 56 c 50 g 46 t

Query Match 49.4%; Score 39; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
DB 105 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 143
|||||

RESULT 14

BH747744 219 bp DNA linear GSS 27-FEB-2002
LOCUS SALK_033710.53.75.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_033710.53.75.x, DNA
sequence.

ACCESSION BH747744
VERSION BH747744
KEYWORDS GS8.
SOURCE GI:18960859
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 219)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)

ACCESSION BH753813
VERSION BH753813
KEYWORDS GS8.
SOURCE GI:18974308
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 215)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..215
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_029677.55.00.x"
/note="lib="Arabidopsis thaliana TDNA insertion lines"
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 63 a 56 c 50 g 46 t

Query Match 49.4%; Score 39; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
DB 105 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 143
|||||

RESULT 14

BH747744 219 bp DNA linear GSS 27-FEB-2002
LOCUS SALK_033710.53.75.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_033710.53.75.x, DNA
sequence.

ACCESSION BH747744
VERSION BH747744
KEYWORDS GS8.
SOURCE GI:18960859
ORGANISM Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 219)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.
Location/Qualifiers
1. 219

FEATURES
source
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_033710.53.75.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT
66 a 57 c 50 g 46 t

Query Match 49.4%; Score 39; DB 17; Length 219;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
DB 109 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 147

RESULT 15
BH747438

LOCUS 220 bp DNA linear GSS 27-FEB-2002
DEFINITION SALK_017469.51.60.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_017469.51.60.x, DNA
sequence.

ACCESSION BH747438
VERSION BH747438
KEYWORDS BH747438.1 GI:18960553
SOURCE GSS.

ORGANISM
thale cress.
Arabidopsis thaliana

Bukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 220)
Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab
, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Ecker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)

JOURNAL Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

COMMENT This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

FEATURES
source
Location/Qualifiers
1. 220

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_017469.51.60.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 67 a 55 c 48 g 50 t

Query Match 49.4%; Score 39; DB 17; Length 220;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
DB 111 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 149

Search completed: May 11, 2003, 04:54:21
Job time: 1272.94 secs

